# Three Essays on Cambodia's Post-Genocide Development

*Sreyphea Chheang*

A Dissertation
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Graduate Program in Economics
York University
Toronto, Ontario

June 2025

## Abstract

This dissertation examines how the Cambodian genocide and agricultural practices have shaped the country's long-run development. It focuses on three outcomes: macroeconomic growth, individual-level human capital, and the gender composition of agricultural labor.

Chapter 1 analyzes the long-term economic effects of the Cambodian genocide using a counterfactual scenario that removes demographic disruptions caused by mass killings. I simulate population structures using fertility and survival rates from 1950 to 2015 and embed them in a production function with heterogeneous labor and intergenerational skill transmission. The results show that although actual GDP per capita initially exceeded the counterfactual due to a higher working-age ratio and more land and capital per worker, this advantage reversed over time. Persistent human capital losses and a delayed recovery in skill composition led to slower productivity growth and reduced long-run economic output.

Chapter 2 uses the sudden, nationwide disruption of the Pol Pot regime (1975 to 1979) as a natural experiment to estimate how being born in an urban area during the genocide affected adult education and wealth. Applying a generalized Difference-in-Differences approach, I use relative district birth size and other indicators to proxy for pre-genocide urbanization. Urban-born cohorts completed 0.02 to 1.6 fewer years of schooling, with the 1977 cohort in Phnom Penh showing the most significant decline in education and wealth. Results are robust across specifications and highlight the lasting human capital impact of forced urban evacuation.

Chapter 3 investigates whether rice cultivation is associated with lower female participation in agriculture. At the farm level, I run OLS regressions of the female labor share on the proportion of land allocated to rice. At the district level, I use both OLS and IV regressions, instrumenting rice yield with elevation based on the suitability of lowland areas for rice production. Results show that farms with more rice land employ fewer women, and districts with higher rice yields have lower female participation. IV estimates confirm a correlation between rice cultivation and reduced female agricultural labor.

Together, these three studies offer insight into how violence and agricultural practices have shaped Cambodia's post-genocide development.

## Dedication

This dissertation is dedicated to my beloved parents, Chheang Chour and Sok Eang, whose faith in me reminded me to believe in myself, even in the most difficult moments. Thank you for your sacrifices and for encouraging my education, even when it was not always expected. Though I could not share this moment with you, this is for both of you. *We did it*!

To my brothers and sister, your constant support and encouragement have meant everything. And to my partner, Sambath Kin, thank you for being my rock, for standing beside me, and for supporting me through every emotional high and low along this milestone Ph.D. journey.

## Acknowledgements

I am deeply and sincerely indebted to my supervisor, Nils-Petter Lagerlöf, for his generous time, invaluable guidance, and unwavering support throughout this journey. His patience, sharp insights, and consistently thoughtful and impressively prompt feedback have shaped not only this dissertation but also my growth as a researcher. I am especially grateful for the countless hours he devoted to discussing my work, encouraging me to refine my thinking, and supporting me through periods of doubt and difficulty. This dissertation, and the person I have become through this process, would not have been possible without his mentorship.

I am also grateful to my committee members, Professors Tasso Adamopoulos and Laura Salisbury, for their insightful suggestions that significantly improved the quality of my research.

Special thanks go to Professor Ahmet Akyol, who kindly served as the chair of my PhD defense and supported me during a critical period in my second year when I was in need of supervision. Without his help, I would not have been connected with my current supervisor.

I would also like to thank Professors Azim Essaji and Janice C.H. Kim for serving as my external examiners and for offering valuable feedback and suggestions to strengthen this work.

Finally, I am grateful to the many ECON 7000 seminar professors who provided helpful comments on my research and presentations. In alphabetical order, I thank Chaoran Chen, Matias Cortes, Antoine Djogbenou, Uros Petronijevic, Georgios Stefanidis, and Pujee Tuvaandorj for their thoughtful feedback and support.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# The Demographic and Economic Effects of the Cambodian Genocide

## 1.1 Introduction

The Cambodian genocide under the Khmer Rouge regime 1975-1979 resulted in the deaths of an estimated 1-3 million people, disproportionately targeting the educated, urban, and middle-class populations, including ethnic minorities (Banister & Johnson, 1993; De Walque, 2006; Heuveline, 1998, 2015; Kiernan, 1996).[1][2] Under the regime, a severe demographic shock led a sharp decline in fertility and a surge in mortality due to execution, exhaustion, starvation, and disease. Famine may have been used as a deliberate instrument of repression, similar to the Soviet use of famine in Ukraine during the 1930s (Markevich et al., 2021).

The genocide triggered a sudden and uneven population decline across age groups, creating a distortion in Cambodia's age structure. Prior to the Khmer Rouge, the Cambodian population was growing at nearly the same rate as neighboring Thailand and Vietnam, despite the U.S. bombing (see Figure 1.1).[3] However, by 2010, the proportion of Cambodians aged 15-49 in 1975 who survived to age 50-80 had declined to nearly half the level observed in the same cohorts in Thailand and Vietnam (see Figure 1.2).

A baby boom emerged immediately around 1979-1980 as a demographic response to excess mortality (Heuveline, 2003), with women aged 25-29 contributing the most (see Figure 1.3). A similar pattern occurred after the 2004 Indian Ocean tsunami, where mothers who lost children were more likely to have another compared to those who had not (Nobles et al., 2015).[4] This

[1] Death for all causes was about 1.67 million in Kiernan (1996), 1.05 million in Banister and Johnson (1993), between 1.2 to 3.4 million in Heuveline (1998), and 1 to 2 million in Heuveline (2015). Papers by Heuveline (1998) and Heuveline (2015) cite various researchers included Kiernan (1996)'s and Banister and Johnson (1993)'s findings of the excess mortality.

[2] The term "Khmer" represents native ethnicity in Cambodia, whereas "Cambodian" refers to all citizens, including ethnic minorities such as Chinese, Vietnamese, and Muslim Chams.

[3] Note that the total population for each country is shown relative to its own 1950 population.

[4] Nobles et al., 2015 find that mothers who lost at least one child are more likely to have another child than childless women and mothers whose children survived. Also, childless women have a higher probability of having a

short-term fertility replacement may also reflect delayed childbearing due to widespread starvation, disease, and malnutrition during the Pol Pot regime.

Figure 1.1: Total population of three countries relative to 1950.

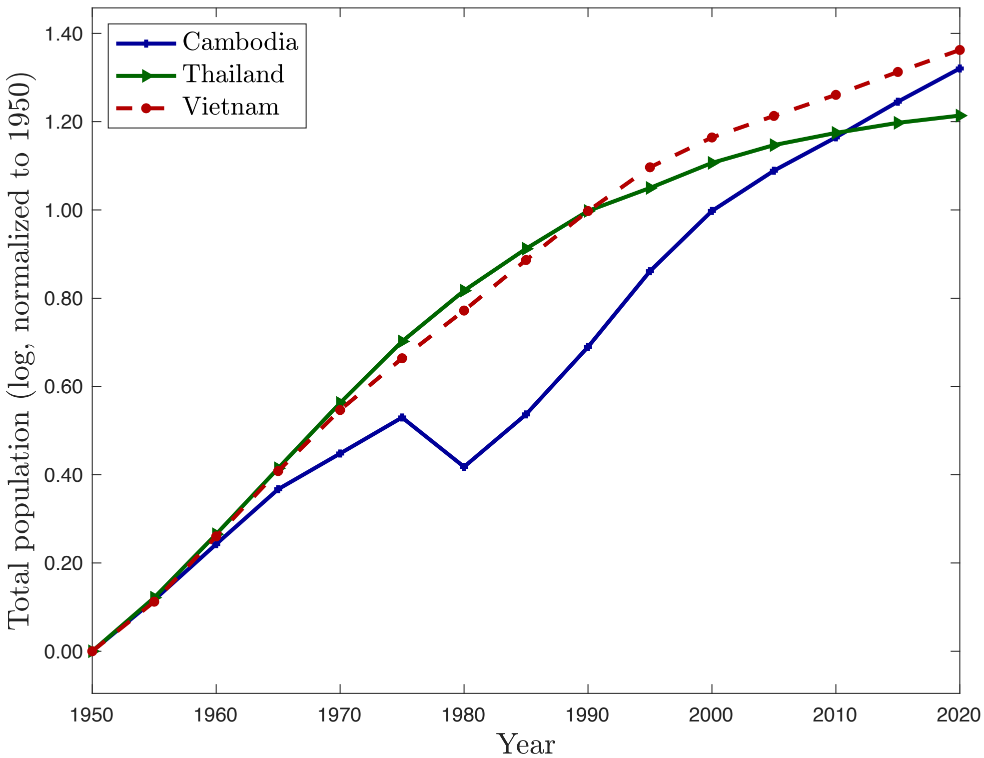
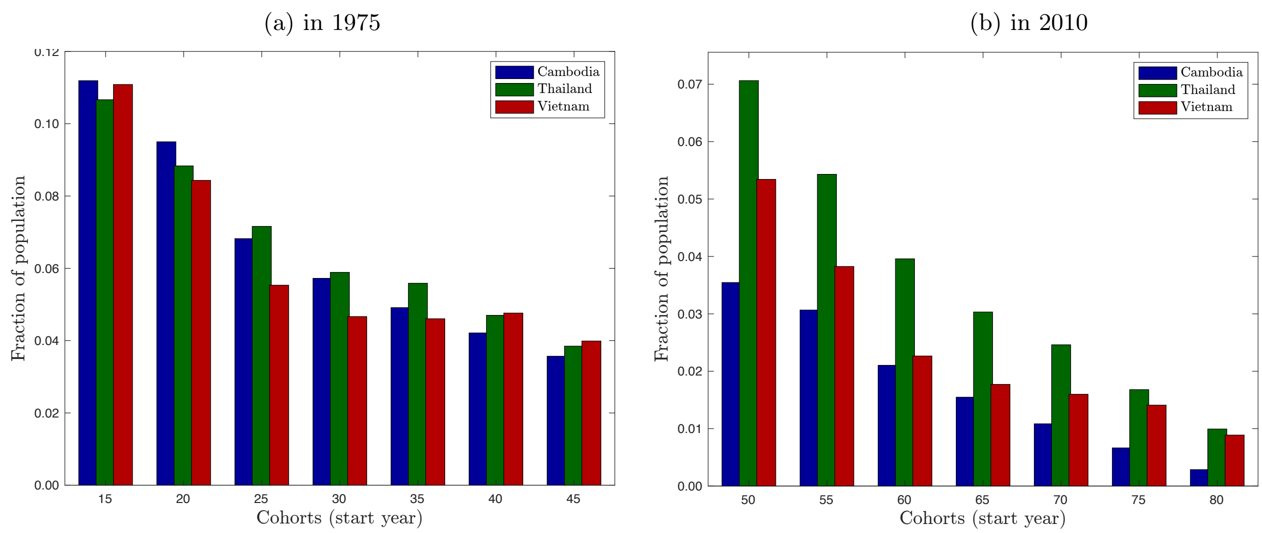



Figure 1.2: Reproductive age groups of three countries.

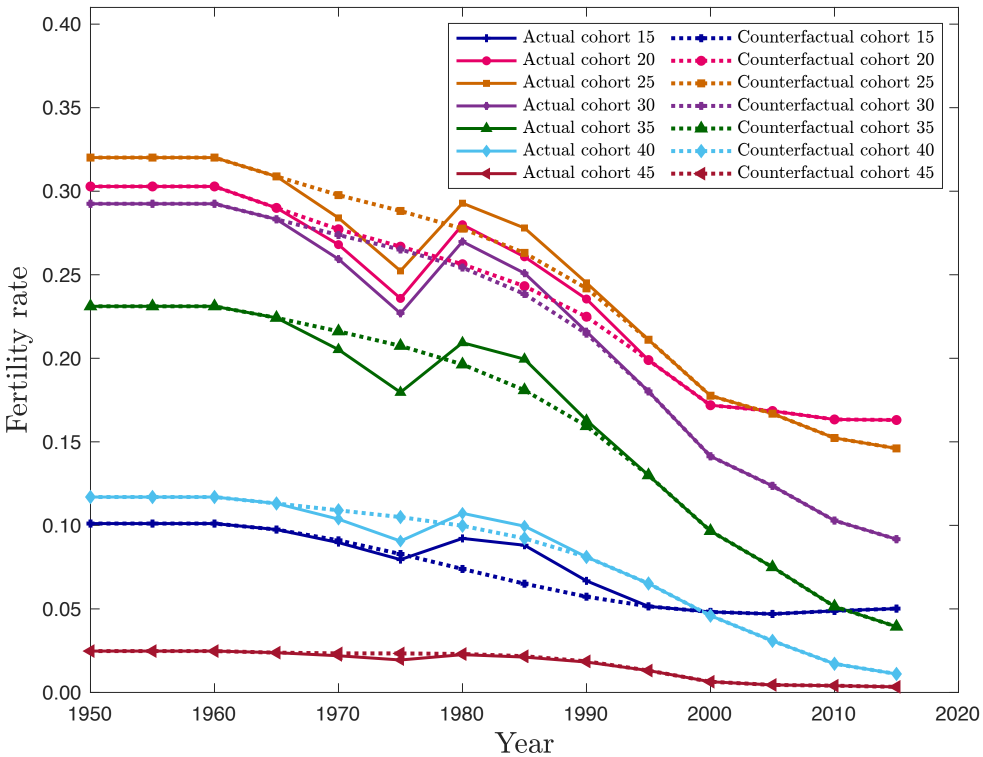



Similar fertility disruptions were observed by Lindstrom and Berhanu (1999) during periods

child than women who did not lose any children. This result contributes to an increase in the overall fertility of 0.7.

of political and economic instability in Ethiopia, highlighting the significant impact of traumatic events on reproductive decision-making.

Figure 1.3: Age-specific fertility.




The mass killings undeniably impacted human capital (De Walque, 2006; Islam et al., 2016), with only 87 out of 1000 scholars and intellectuals surviving the regime (Clayton, 1998). Similarly, there is limited research on how the resulting shifts in age structure have affected Cambodia's long-term economic performance. Following a sharp rise in fertility rates immediately after the genocide, fertility began to decline after 1985 (see Figure 1.3), while survival rates continued to rise (see Figure A.15). The genocide also reshaped political attitudes, increasing support for pluralism among survivors and subsequent generations (Bühler & Madestam, 2024), and contributed to long-term economic loss similar those observed after the Holocaust Acemoglu et al. (2011).[5]

This paper investigates how changes in age structure and skill composition explain the long-run demographic and economic consequences of the Cambodian genocide. Shifts in age distribution affects productivity and output (Feyrer, 2007), as the economy benefits from with a larger share of the population enters the labor force. For instance, the baby boom cohort's entry into the the labor market raises the working-age to non-working-age population ratio, generating a "demographic dividend" (Bloom & Canning, 2008; Bloom et al., 2011).[6] Similarly, Zélity (2023) shows that

[5] Acemoglu et al. (2011) find the Holocaust caused a long-term economic loss and adverse effects on the social and political structure.

[6] Feyrer (2007) emphasizes that Japan had high productivity growth between 1960 and 1980 was largely driven by its demographic advantage, particularly a higher proportion of workers in their 40s. In contrast, the United States had

balanced cohort sizes can enhance productivity through complementary skills across age groups.

To study these dynamics, I first construct counterfactual population time paths using age-specific fertility and survival data from 1950-2015, simulating Cambodia's demographic structure in the absence of the genocide (see Section 1.4 for construction details). This allows for direct comparison with observed demographic trends. I then estimate the skill composition of the labor force and aggregate output using a Constant Elasticity of Substitution (CES) production function, applied to both actual and counterfactual population and age structures within a Solow-type framework. The model incorporates skill-heterogeneous labor, capital, and land to capture the economic implications of demographic change. This enables a comparative analysis of the actual economy impacted by the genocide and a counterfactual scenario without it.

One challenge in this paper is the estimation of age-earning profiles, as such data for Cambodia is either not publicly available or extremely difficult to obtain over the full time horizon of interest. To address this, a cohort-wage profile is constructed using wage and income data from the 2010 Cambodia Socio-Economic Survey (CSES), as provided by Humphreys (2015). Although the estimated profile may not fully reflect actual wage dynamics, its hump-shaped pattern is consistent with the commonly assumed concave lifetime age-earning profile. Based on this, cohort-specific technology parameters and the productivity share of high-skilled labor are derived from the estimated wage profile.

The results show that excess mortality and displacement during the genocide caused a significant divergence between actual and counterfactual population and age structures after the 1980s. The larger this divergence, the more persistent the effects of the genocide appear to be. Indeed, the demographic impact of the genocide remains evident to the present day. The observed shortfall in population in 1980 implies that there were approximately 1.8 million excess deaths when compared to a no genocide scenario. This figure is consistent with estimates by Heuveline (1998, 2015). The smaller size of the 35-39 cohort in 2010 indicates a severe fertility decline during the Pol Pot regime. This finding corresponds to Banister and Johnson (1993) estimate about $216,000$ fewer births in 1979. In contrast, the difference in the 30-34 cohort reflects a post-genocide fertility replacement. Indeed, the genocide caused an uneven age distribution, which continues to affect the working-to-non-working-age population ratio, primarily through the big volume of the baby boom cohorts.[7]

GDP per capita was higher in the actual scenario compared to the counterfactual between 1970-1990, but lower from 1995 onward. This pattern can be explained by a higher working-to-non-working-age ratio, as discussed in Ashraf et al. (2013), Bloom and Canning (2008), Bloom et al. (2011). The genocide likely caused disproportionate deaths among the young and elderly, resulting in a larger share of the working-age population. According to Meng-Try (1981) about $30,000$ young children died in the last six months of 1975 due to malnutrition and poor sanitation. However, the counterfactual economy grew more substantially, as it did not experience mass killings

fewer workers in this age range and experienced slower productivity growth. These differences reflect each country's distinct historical birth rates and, consequently, their differing age structures.

[7] The ratio of the working-to-non-working-age population is the division between the number of the working-age population (ages 15-69) and the number of the dependent population (ages 0-14 and 70+).

and the demographic distortions of the post-genocide baby boom. Moreover, The skill composition effect shows that the actual economy was worse than the counterfactual in 2010. One plausible explanation is the intergenerational persistence of skill formation, which takes longer to recover to pre-genocide levels. The stock of high-skilled children depends on the educational background of their parents. Because many educated individuals perished during the Pol Pot regime, the effects on skill composition have persisted over time.

This paper contributes to the literature by examining how the Cambodian genocide reshaped the country's age structure and how these changes affected long-term economic growth. While previous research links age structure to growth through the demographic dividend, emphasizing the role of rising working age populations and declining fertility rates (Ashraf et al., 2013; Bloom et al., 2011), few studies explore how mass violence distorts this channel.[8] I focus on how changes in the working-age population ratio and intergenerational skill composition influence output and productivity. Using a counterfactual simulation, I compare actual demographic and economic outcomes to those in a scenario without genocide. This approach isolates both demographic and skill composition effects and shows how genocide alters labor force structure and shaped the demographic dividend after 1995, when Cambodia began a standard demographic transition.

This paper also contributes to the ongoing literature on the long-term demographic and economic effects of civil conflict and genocide. Previous studies show that poverty is closely linked to war and conflict, which are more likely to occur in the least developed and least democratic countries, particularly where natural resources are abundant, governance is weak, and opportunity costs are low (Blattman & Miguel, 2010; Collier, 2008; Collier & Hoeffler, 2004; Easterly et al., 2006). War and violence cause the short-term disruptions to population (Davis & Weinstein, 2002), human capital, and physical capital (Miguel & Roland, 2011).[9] However, a growing body of research highlights that these shocks can also have long term consequences. Acemoglu et al. (2011) find that mass killings during the Holocaust had lasting adverse effects on economic and social structures. Riano and Valencia Caicedo (2024) show that U.S. bombing in Laos led to declines in educational attainment and slower economic growth decades later.

In the context of the Cambodian genocide, Merrouche (2011) finds that landmine exposure led to an average loss of half a year of schooling for individuals who were not of school age before the mass killing. Islam et al. (2016) show that the genocide affected both educational attainment and earnings, with men receiving 0.9-1.1 fewer years of schooling and women 0.6-0.9 fewer years. Additional evidence from De Walque (2006) attributes long term declines in education to the destruction of schools and the systematic targeting of urban, educated men. Similar findings are reported in De Walque and Verwimp (2010) for Rwanda.

The paper proceeds as follows. Section 1.2 provides a brief historical of Cambodian genocide.

[8] According to Bloom and Canning (2008), Bloom et al. (2011), the demographic dividend is referred to the change in the age structure of the population that generates a faster economic growth.

[9] Davis and Weinstein (2002) find that bombed cities in Japan eventually returned to their prewar population growth rates. Similarly, Miguel and Roland (2011) report no long-term impacts on either physical or human capital 25 years after the U.S. bombing in Vietnam.

Section 1.3 describes the data sources used to construct population inputs. Section 1.4 outlines the model and reconstructs population by age and skill. Section 1.5 presents the initial conditions and parameter calibration. Section 1.6 compares simulated actual and counterfactual demographic and economic outcomes. Section 1.7 concludes.

## 1.2 Background

A few years before the Pol Pot regime, Cambodians experienced two catastrophic events. First, over half a million tons of bombs were dropped on the countryside by the U.S. secret mission, resulting in the deaths of more than 100,000 civilians (Kiernan, 2004). Second, in 1970, a coup led by General Lon Nol and Sisowath Sirik Matak overthrew Prince Norodom Sihanouk just days after he departed for a diplomatic visit to China and the Soviet Union.[10] Although the new regime initially claimed victory, it quickly began to unravel. By April 1975, the Khmer Rouge had ousted Lon Nol's government and seized control of the country under a radical Communist system.

The Khmer Rouge leader, known as Pol Pot, was born as Saloth Sar into a farming family in Kompong Thom province. He later received a scholarship to study radio-electricity in Paris, France, but did not complete his degree (Kiernan, 1996).[11] Upon returning to Cambodia, he became the leader of the Communist Party of Kampuchea. The regime's seizure of power in 1975 was driven by a radical vision of a Utopian society, one without class divisions, wealth disparities, or private property, where all citizens would live under the same standard.

Phnom Penh was captured by the Khmer Rouge on April 17, 1975. Immediately afterward, the entire population was forced to evacuate to rural villages across six administrative zones, under the false promise that they would return within a few days.[12] Yet, the displacement lasted three years, eight months, and twenty days. During the chaotic evacuation, families were separated in the overcrowded streets as people scattered in all directions. According to Chandler (2018) and Kiernan (1996), approximately two million people had been living in the city at the time, and about 10,600 died during the forced march out of the capital (Kiernan, 1996).[13] While many attempted to flee the country and seek refuge in Thailand or Vietnam, emigration during the Khmer Rouge regime was nearly impossible. If at all possible, one was unlikely to survive the journey or reach the destination.

Once city dwellers arrived at their destinations, they were each given a small plot of empty land and a limited portion of rice. Those who had been relocated were labeled as "New People" or "17th April people," while residents already living in the villages were referred to as "Base People" or "Villagers." After six to eight months, many of the New People were again forced to move to different locations.[14] Both groups, however, were subjected to long hours of manual labor, assigned

[10] Sisowath Sirik Matak is king Sihanouk's cousin.

[11] Kiernan (1996) and Chandler (2018) are among the well-written books about the Pol Pot, and civil war before the Pol Pot regime.

[12] The six zones are North, Northeast, Northwest, West, Southwest, and East, while Phnom Penh is the center zone and the Pol Pot's base.

[13] The toll number depends on author interviewed 36 people out of a hundred or more people about the evacuation; page 48.

[14] Ebihara and Kiernan (1993) has a well written about Base people's experience under the Khmer Rouge and post

strict quotas and tasks, yet received only minimal rice allowance. This combination of overwork and starvation was devastating, even for Base People who were already accustomed to physically demanding labor as peasants (Ebihara & Kiernan, 1993). As a result, deaths from exhaustion and famine contributed significantly to the overall death toll, in addition to executions and other violent causes.

School-age children were taught using a curriculum written by the Khmer Rouge, while elderly people were assigned to care for infants. However, they received the same minimal rice allowances as other adults and faced a high risk of dying from famine and malnutrition. According to Meng-Try (1981), approximately $30,000$ young children died during the last six months of 1975 alone, due to malnutrition and unhygienic conditions. However, not all survivors experienced famine in the same way, as it depended on where they lived and who the local Khmer Rouge cadres were.

January 7, 1979, marked the end of the Pol Pot regime, which was overthrown by a resistance group of former Khmer Rouge members, including Heng Samrin, Hun Sen, and Chea Sim, with the support of Vietnamese troops. This event brought an end to the period of mass killings, starvation, and famine. Between 1975 and 1979, Kiernan (1996) estimates that approximately $879,000$ of the so-called "New People" and $792,000$ of the "Base People" died out of total populations of 3 million and 4.8 million, respectively. As a result, Cambodia's total population in 1980 was estimated at around 6.5 million (Banister & Johnson, 1993) and 6.8 million (Neupert & Prum, 2005).

The new government, known as the People's Republic of Kampuchea, was initially led by three former Khmer Rouge officials. In the late 1980s, Hun Sen was appointed prime minister, a position he has held continuously since then. Cambodia held its first general election in 1993, organized by the United Nations Transitional Authority in Cambodia (UNTAC), marking a significant step toward political reconstruction.

## 1.3 Data

This section outlines the data used to construct the counterfactual age distribution, age-specific migration factors, and total population time paths. The data, covering the periods 1950-2020, are extracted from the United Nations Population Division (UNDP) (2019a). According to UNDP, demographic indicators during the Khmer Rouge period, including the population, survival and mortality rates, and fertility rates, are based on population reconstructions for the years 1962 to 1998.[15] For other years, the data are estimated based on census and survey data from available years such as 1962, 1998, and 2008 Censuses, the 1996 Demographic Survey (DS), and the 2004 and 2013 Inter-censal Population Survey (ICPS).

The primary source of data is the population distribution by five broad age groups and both sexes, reported as annual mid-year estimates covering ages 0 to 100 and older. The fraction of each female cohort is calculated by dividing the number of females in a given age group by the total

Khmer Rouge regime.

[15] Further specific details can found at: https://population.un.org/wpp/Download/Metadata/Documentation/.

population of that cohort. Moreover, instead of using mid-year population estimates, I adjust the data to represent the population as of January 1st of each calendar year. This is done by averaging the values from two consecutive years, except for 1950. For example, the population in 1955 is calculated as the average value of the 1954 and 1955.

Age-specific survival probabilities are taken directly from the Abridged Life Tables (ALT). These probabilities are provided for age intervals beginning with ages 0-1 and 1-4, followed by five-year groups from 5-9, 10-14, and up to 95-99. Since the population data include both sexes, survival rates are assumed to be the same for males and females within each age cohort for simplicity.

Age-specific fertility rates represent the number of births by five-year age groups for women between ages 15 and 49. Each observation corresponds to a five-year interval from 1950 to 2015. The data are estimated using the method of proportionate age-specific fertility rates, based on birth-history data from the 1998 Demographic and Health Survey and National Survey (DHS-NS), as well as the 2000, 2005, 2010, 2014 DHS. However, the UNDP does not specify how or where fertility rates were estimated for the years 1950 to 1965, and the reported values remain constant across those years.

## 1.4 Setting

This section outlines the quantitative framework used to analyze the long-run economic effects of the Cambodian genocide. The model integrates simulated demographic dynamics into a production function featuring capital, land, and labor differentiated by skill type. Actual and counterfactual population paths are generated using a cohort-based demographic model. Output is determined through a nested CES production function that captures imperfect substitution across both age groups and skill types. The model captures changes in labor composition, intergenerational skill transmission, and capital accumulation, allowing for dynamic comparison between actual and counterfactual scenarios.

### 1.4.1 Production function and skilled heterogeneity labor

This paper uses a Constant Elasticity of Substitution (CES) production function with inputs of labor, capital, and land. The economy is closed, produces a single good, and does not experience technological progress or exogenous total factor productivity (TFP) growth. This assumption follows Ashraf et al. (2013), who argue that in many developing countries, technology is largely imported and thus independent of domestic demographic dynamics. In this setting, income is generated from labor (disaggregated by skill type), land, and physical capital, with each component contributing to capital accumulation according to its respective saving rate.

The total output at time $t$ is defined as

$$Y_t = K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta}, \tag{1.1}$$

where $\alpha$ is a share of capital, and $\beta$ is a share of the potential labor. $1 - \alpha - \beta$ is a share of the land, where $\alpha + \beta < 1$. $K_t$ is the capital stock, $L_t$ is effective labor input (explained further below), and $X_t$ is productivity-augmented land, reflecting improvements in land quality and usage efficiency over time. For simplicity, $X_{1950}$ is normalized to 1.

#### 1.4.1.1 Effective aggregate labor

Aggregate labor input $L_t$ is constructed using a nested CES function that captures heterogeneity in both skill level and age-based productivity. The outer layer aggregates labor across age cohorts, with each group weighted by its age-specific productivity to reflect life-cycle differences in efficiency. The inner layer reflects the imperfect substitutability between high-killed and low-skilled workers within each age cohort. This nested structure enables the model to trace how both demographic shifts and changes in the skill distribution determines aggregate output over time. In each period $t$, each age group $i$ is assumed to supply one unit of labor.

The outer layer aggregates labor across all age cohorts in period $t$, each age group $i$ supply one unit of labor

$$L_t \quad = \quad \left[ \sum_{i=0}^{95} \gamma_i \widetilde{P}_{i,t}^{\rho} \right]^{\frac{1}{\rho}}, \tag{1.2}$$

where $\gamma_i$ represents the productivity weight for each age cohort and captures the relative efficiency of individuals over the life cycle. Note that $\sum_{i=15}^{65} \gamma_i = 1$ for $i \in \{15, 20, ..., 65\}$, and $\gamma_i = 0$ otherwise. The working-age population is defined from 15 to 69. The parameter $\rho$ captures the substitutability between labor from different age groups.

Assuming $\rho < 1$ reflects imperfect substitution, consistent with the idea that younger workers provide greater physical strength, while older workers contribute accumulated experience. These differences imply that labor inputs from different age cohorts are more complementary than interchangeable, and thus cannot be easily substituted for one another in production. In contrast, the standard assumption of $\rho = 1$ implies perfect substitutability across age groups. In this case, the Equation (1.2) becomes linear, and the productivity weights $\gamma_i$ reduce to constant values across cohorts. This simplification treats labor from all age groups as equally productive on a per-unit basis, regardless of age-related differences in experience or capacity, and ignores potential distortions arising from demographic imbalances. By setting $\rho < 1$, the model explicitly incorporates age-related heterogeneity in labor productivity and allows demographic structure to influence aggregate output in a non-linear way.

The inner layer aggregates labor inputs within each cohort by combining high-skilled and low-skilled workers

$$\widetilde{P}_{i,t} = \left[ (1 - \lambda) \left( P_{i,t}^{L} \right)^{\eta} + \lambda \left( P_{i,t}^{H} \right)^{\eta} \right]^{\frac{1}{\eta}}, \tag{1.3}$$

where $P_{i,t}^{L}$ and $P_{i,t}^{H}$ denote low-skilled and high-skilled labor, respectively. The parameter $\lambda \in [0, 1]$ captures the relative productivity of high-skilled labor compared to low-skilled labor and is assumed to be time-invariant.

The parameter $\eta$ governs the substitutability between high-skilled and low-skilled labor within each age cohort and is assumed to satisfy $\eta < 1$. This assumption reflects the idea that high-skilled and low-skilled workers perform distinct roles in production and are not easily interchangeable. A lower value of $\eta$ implies that these two labor types are more complementary than substitutable, highlighting the importance of skill diversity within cohorts.

When $\eta \to 1$, the Equation (1.3) converges to a linear function, and the two skill types become perfect substitutes. When $\eta = 0$, the function becomes Cobb-Douglas form, implying unit elasticity of substitution. As $\eta \to -\infty$, the Equation (1.3) converges to a Leontief form, where high- and low-skilled labor are perfect complements, and effective labor is constrained by the scarcest input.

By assuming $\eta < 1$, the model captures the idea that a balanced mix of skills enhances productivity, and that losses in one skill group cannot be easily offset by increases in the other. This specification assumes constant skill productivity over time and does not account for changes in educational systems or the broader evolution of human capital.

The expression for $\widetilde{P}_{i,t}$ can also be rewritten in terms of the high-skilled share of cohort $i$ at time $t$, denoted $\theta_{i,t} = \frac{P^H_{i,t}}{P_{i,t}}$, where $P_{i,t}$ is the total size of the cohort $i$ (explained further below). This gives

$$\begin{aligned} \widetilde{P}_{i,t} &= \left[(1-\lambda)(1-\theta_{i,t})^{\eta} + \lambda\theta_{i,t}^{\eta}\right]^{\frac{1}{\eta}} P_{i,t} \\ &= \Phi_{i,t} P_{i,t}. \end{aligned} \tag{1.4}$$

where $\Phi_{i,t} \in [0,1]$ and $\Phi_{i,t} = (1-\lambda)(1-\theta_{i,t})^{\eta} + \lambda\theta_{i,t}^{\eta}$. $\Phi_{i,t}$ is a skill adjustment factor that scales the effective size of cohort $i$ in time $t$ based on its internal skill composition. In other words, it acts as a productivity-weighted multiplier that increases the cohort's contribution to aggregate labor input when high-skilled workers make up a larger share of the cohort.

The expression for total effective labor input in Equation (1.2) can now be rewritten as

$$L_t = \left[\sum_{i=0}^{95} \gamma_i \Phi_{i,t}^{\rho} P_{i,t}^{\rho}\right]^{\frac{1}{\rho}}. \tag{1.5}$$

In this form, aggregate labor input depends on cohort size $P_{i,t}$, the productivity weight $\gamma_{i,t}$, and the skill composition within each cohort through $\Phi_{i,t}$. This structure allows both demographic changes (such as age structure and cohort size) and human capital distribution (skill shares within cohorts) to determine output through their effects on effective labor supply.

#### 1.4.1.2 Factor prices

This section defines the prices of the three primary production inputs: labor, capital, and land. The wage rate $w_t$ represents the payment to effective labor. It depends on both the size of the labor force and the productivity of workers, which varies by age and skill type. The return to capital $r_t$ and the return to land $q_t$ reflect the rental prices of their respective inputs. All factor prices are derived from the production function in Equation (1.1), assuming firms maximize profits in competitive markets.

Since the labor force includes both high-skilled and low-skilled workers, each type earns a different wage. Let $w_{i,t}^j$ denote the wage earned by a worker of type $j \in \{H, L\}$ in cohort $i$ at time $t$. The wage for high-skilled labor is higher in more productive cohorts due to their greater value-added. Wages are derived by differentiating the production function with respect to $P_{i,t}^j$

$$w_{i,t}^H = \lambda \theta_{i,t}^{\eta-1} \Phi_{i,t}^{\rho-\eta} \widetilde{w}_{i,t}, \tag{1.6}$$

$$w_{i,t}^L = (1-\lambda)\left(1-\theta_{i,t}\right)^{\eta-1} \Phi_{i,t}^{\rho-\eta} \widetilde{w}_{i,t}, \tag{1.7}$$

where $\widetilde{w}_{i,t} = \beta \gamma_i \frac{Y_t}{L_t} \left(\frac{L_t}{P_{i,t}}\right)^{1-\rho}$. The term $\widetilde{w}_{i,t}$ reflects the baseline wage for cohort $i$ in the absence of skill heterogeneity. It is determined by the cohort's productivity weight, its relative size in the labor force, and the overall labor-output ratio. In other words, it reflects what the cohort would earn based on age-specific productivity and labor share, excluding the effects of skill composition.

When the share of high-skilled labor in a cohort $\theta_{i,t}$ increases, wages respond asymmetrically across skill types due to substitution and productivity dynamics in the nested CES structure. For high-skilled workers, the term $\theta_{i,t}^{\eta-1}$ decreases when $\eta < 1$ due to diminishing returns effect. Although a higher $\theta_{i,t}$ also raises the cohort's overall productivity $\Phi_{i,t}^{\rho-\eta}$, this positive effect is typically weaker than the diminishing marginal returns when $\rho \leq \eta$. As a result, the wage effect for high-skilled labor is typically negative at high levels of $\theta_{i,t}$. Conversely, low-skilled wages rise with $\theta_{i,t}$ because the declining share of low-skilled labor $(1-\theta_{i,t})$ increases their marginal value through the term $(1-\theta_{i,t})^{\eta-1}$.

Moreover, the rental price of capital is

$$r_t = \alpha \frac{Y_t}{K_t}, \tag{1.8}$$

and the rental price of land is

$$q_t = (1-\alpha-\beta) \frac{Y_t}{X_t}. \tag{1.9}$$

These expressions derive from the marginal products under constant returns to scale, where capital, labor, and land earn fixed shares $\alpha$, $\beta$, and $1-\alpha-\beta$, respectively. Changes in demographic structure and skill composition alter effective labor input, thereby affecting income distribution across factors.

#### 1.4.1.3 Per-capita income

A change in the population structure caused by mass killings directly affects the standard of living over time. In this framework, changes in output per capita resulting from demographic shocks can be evaluated through multiple channels. Specifically, gross domestic product per capita can be decomposed into three structural components: the ratio of capital to labor ($\frac{K_t}{L_t}$), the ratio of land to labor ($\frac{X_t}{L_t}$), and the ratio of labor to total population ($\frac{L_t}{P_t}$).

To derive this decomposition, particularly when comparing actual and counterfactual scenarios, the production function in Equation (1.1) can be rearrange as

$$y_t = \frac{K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta}}{P_t} = \frac{K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta}}{P_t} \left( \frac{L_t^{1-\alpha}}{L_t^{1-\alpha}} \right) = k_t^{\alpha} x_t^{1-\alpha-\beta} l_t, \tag{1.10}$$

where $k_t = \frac{K_t}{L_t}$ is capital per worker, $x_t = \frac{X_t}{L_t}$ is land per worker, and $l_t = \frac{L_t}{P_t}$ is labor-population ratio. The labor-to-population ratio $l_t$ reflects both the age distribution and skill composition of the population, capturing the demographic factors that shape labor supply. This decomposition enables identify how each component contributes to changes in living standards following a population shock.

A lower labor-to-population ratio reduces effective labor supply and, in turn, income per capita, even if capital and land remain constant. Although capital per worker may temporarily increase due to a smaller labor force, this does not necessarily translate into higher living standards. Long term recovery depends on rebuilding both the population size and the productivity of the working age population. This framework provides a basis for evaluating how demographic and structural factors affect actual and counterfactual economic outcomes.

### 1.4.2 Population dynamics

#### 1.4.2.1 New-born population

The size of the newborn population in period $t+5$, denoted $P_{0,t+5}$ is determined by births conceived in period $t$. The variable $n_{i,t}$ represents the age-specific fertility rate, defined as the average number of births per woman per year in age group $i \in [15, 49]$. This is calculated using a Total Fertility Rate (TFR) approach under the assumption of no direct fertility shocks.

The number of newborn age-0 is given by

$$P_{0,t+5} = 5 \left[ \sum_{i=15}^{45} s_{0\text{-}1,t} n_{i,t} m_{i,t} f_{i,t} P_{i,t} \right], \tag{1.11}$$

where $f_{i,t}$ is the share of females in the population for age group $i$. $P_{i,t}$ is the total population in cohort $i$.[16] $s_{0\text{-}1,t}$ is the probability that a newborn survives to age 1 and captures infant mortality. $m_{i,t}$ is the migration factor (explained further below) and accounts for net migration of women, including the movement of newborns with their mothers.

The multiplication by 5 accounts for the five-year width of each age cohort, since women in each cohort are assumed to have a constant annual fertility rate $n_{i,t}$ over that interval. The use of the time index $t+5$ on the left-hand side of Equation (1.11) follows standard demographic conventions for updating total fertility rate series in five-year intervals (see Shryock et al., 1975, pp. 287-288 for

[16] Note that the total number of females in age group $i$ is given by the product $f_{i,t}$ and $P_{i,t}$.

details). For example, children conceived in year $t$ are born and counted as ages 0-4 in year $t+5$. Specifically, births in 1950 are observed as age-0 cohorts in 1955.

To account for infant survival and migration during early childhood, the age-5 population in period $t+5$ is calculated as

$$P_{5,t+5} \quad = \quad s_{\text{1-4},t} m_{0,t} P_{0,t}, \tag{1.12}$$

where $s_{\text{1-4},t}$ is the survival rate from ages 1 to 4, and $m_{0,t}$ is the migration adjustment for children aged 0-4. To avoid over-adjusting for infant mortality, Equation (1.12) uses $s_{\text{1-4},t}$ rather than a combined survival rate $s_{\text{0-4},t}$. Although one might consider merging Equations (1.11) and (1.12) into a single expression using $s_{\text{0-4},t}$, this would introduce temporal inconsistency. This is because the product of $s_{\text{0-1},t}$ and $s_{\text{1-4},t}$ draw from different time indices and therefore does not equal to $s_{\text{0-4},t}$. For example, the $P_{0,1955}$ depends on infant survival from 1950, specifically $s_{\text{0-1},1950}$, while $P_{5,1960}$ depends on childhood survival from 1955, $s_{\text{1-4},1955}$. Using a unified term $s_{\text{0-4},1950}$ would ignore this time lag and inaccurately represent the actual survival path of the cohort.

#### 1.4.2.2 Cohort size

Population levels evolve over time through the effects of survival and migration. These factors affect not only the total size of the population but also its age structure. This mechanism captures how genocide-related mortality and displacement altered that structure by sharply reducing the size of specific cohorts.

The evolution of each age cohort is captured by the following cohort transition equation

$$P_{i+5,t+5} \quad = \quad s_{i,t} m_{i,t} P_{i,t}, \tag{1.13}$$

where $i \in \{0, 5, ..., 95\}$ represents the starting age of each 5-year age cohort, and $t \in \{1950, 1955, ..., 2015\}$ denotes the first year of each 5-year period. $P_{i,t}$ is the size of cohort $i$ in period $t$, and $P_{i+5,t+5}$ is the size of the same cohort for the next period. The parameter $s_{i,t}$ is the age-specific survival rate, capturing the proportion of individuals in cohort $i$ who survive to the next period. This includes mortality from all causes, such as disease, malnutrition, or accidents. The term $m_{i,t}$ is the age-specific migration factors and captures the net effect of in-migration and out-migration.

The left-hand side of Equation (1.13) decreases when survival probabilities are low or when there is substantial out-migration. As a result, cohort sizes shrink in the following period. This dynamic was evident during the genocide period, when survival rates fell sharply due to starvation, disease, and conflict-related mortality. Simultaneously, mass displacement and forced migration further reduced cohort sizes. These shocks are captured in the model through declining survival parameters and migration factors. As a result, cohort sizes decline significantly in subsequent periods.

#### 1.4.2.3 Newborn by skill type

Building on the skill-specific population dynamics, the composition of each newborn cohort is divided into high-skilled and low-skilled individuals. This division reflects intergenerational skill transmission, where the likelihood of a child being high-skilled depends on the skill type of their parents.

Both high-skilled and low-skilled parents can have children of either skill type, though the probability differ. Let $\pi^H \in [0, 1]$ denote the probability that high-skilled parents have high-skilled children, and $\pi^L \in [0, 1]$ denote the corresponding probability for a low-skilled parent. It is assumed that $\pi^H > \pi^L$, meaning high-skilled parents are more likely to transmit their skills to the next generation.

The total number of newborns in period $t + 5$ is sum of high-skilled and low-skilled children

$$P_{0,t+5} \quad = \quad P^H_{0,t+5} + P^L_{0,t+5}. \tag{1.14}$$

The number of high-skilled newborns is

$$P^H_{0,t+5} \quad = \quad 5\left(\sum_{i=15}^{45} s^H_{0\text{-}1,t} n_{i,t} \pi^H m_{i,t} f_{i,t} P^H_{i,t} + \sum_{i=15}^{45} s^L_{0\text{-}1,t} n_{i,t} \pi^L m_{i,t} f_{i,t} P^L_{i,t}\right). \tag{1.15}$$

Similarly, the number of low-skilled newborns is

$$P^L_{0,t+5} \quad = \quad 5\left(\sum_{i=15}^{45} s^H_{0\text{-}1,t} n_{i,t}(1-\pi^H) m_{i,t} f_{i,t} P^H_{i,t} + \sum_{i=15}^{45} s^L_{0\text{-}1,t} n_{i,t}(1-\pi^L) m_{i,t} f_{i,t} P^L_{i,t}\right), \tag{1.16}$$

where $s^H_{0\text{-}1,t}$ and $s^L_{0\text{-}1,t}$ represent the newborn survival rates from birth to age 1 for children of high-skilled and low-skilled parents, respectively. The parameters $n_{i,t}$, $m_{i,t}$, and $f_{i,t}$ refer to the age-specific fertility rate, migration rate, and share of females, respectively, as previously defined in Section 1.4.2.1. These are assumed to be uniform across skill types for simplicity.

Equation (1.15) defines the total number of high-skilled newborns as the sum of births from high-skilled and low-skilled parents. The first term captures the births to high-skilled parents with probability $\pi^H$, while the second term accounts for high-skilled children born to low-skilled parents with probability $\pi^L$. Equation (1.16) follows a similar structure for low-skilled newborns.

Under the Khmer Rouge regime, the entire population faced high mortality risks. However, individuals with higher education and urban backgrounds were disproportionately affected, facing greater risks of execution, starvation, and disease. To accurately capture this severe mortality shock, the model introduces a one-time, cohort-specific parameter $\psi_{i,t}$. This parameter adjusts survival rates for selected age groups to reflect the regime's disproportionate impact on specific segments of the population. In particular, $\psi_{i,t}$ is applied to high-skilled cohorts, reflecting their increased vulnerability during this period (explained further below).

In the absence of direct data on survival rates by skill type, this mortality shock is imposed as a fixed exogenous parameter. The survival rate for high-skilled newborns is guessed as

$$s^H_{0\text{-}1,t} \quad = \quad (1+\psi_{0,t})\, s_{0\text{-}1,t}, \tag{1.17}$$

where $\psi_{0,t} \in \left[-1, \frac{1}{s_{0\text{-}1,t}} - 1\right]$ is an exogenous, time-dependent parameter that reflects a survival penalty for high-skilled newborns. This guess ensures that both $s^H_{0\text{-}1,t}$ and $s^L_{0\text{-}1,t}$ remain bounded within the interval $[0,1]$. The survival rate $s_{0\text{-}1,t}$ is newborn overall survival rate as discuss in Section (1.4.2.1).

Using equations (1.14) through (1.17), the survival rate for low-skilled newborns is derived as

$$s^L_{0\text{-}1,t} \quad = \quad \frac{1-(1+\psi_{0,t})\,\tau_t}{1-\tau_t} s_{0\text{-}1,t}, \tag{1.18}$$

where

$$\tau_t = \frac{\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P^H_{i,t}}{\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}}, \tag{1.19}$$

represents the share of high-skilled births from high-skilled parents. In other word, $\tau_t$ captures parental skill position at birth.

Therefore, the total survival rate for the newborn cohort is given by

$$s_{0\text{-}1,t} \quad = \quad \tau s^H_{0,t} + (1-\tau)\, s^L_{0,t}. \tag{1.20}$$

Additional details on the derivation and calibration of these expressions are provided in the Chapter 1 Appendix.

#### 1.4.2.4 Skilled-specific population dynamic

The population is categorized by both age cohort and skill type, distinguishing between high-skilled $P^H_{i,t}$ and low-skilled $P^L_{i,t}$ individuals in each cohort $i$ and time $t$. The total population can thus be expressed as

$$P_{i,t} \quad = \quad P^H_{i,t} + P^L_{i,t}. \tag{1.21}$$

The population dynamics for each skill type are defined by

$$P^j_{i+5,t+5} \quad = \quad s^j_{i,t} m_{i,t} P^j_{i,t}, \tag{1.22}$$

where $j \in \{H, L\}$ represents high-skilled or low-skilled type. The term $s^j_{i,t}$ is the skill-specific survival rate for type $j$, while $m_{i,t}$ is the migration factor. Migration is assumed to be the same across skill types and gender, and constant from birth onward.

The overall survival rate $s_{i,t}$ is obtained by applying Equation (1.21) one cohort and period forward to $i+5$ and $t+5$, respectively. Using Equation (1.22), the population at $t+5$ can be written as the sum of high-skilled and low-skilled components. Substituting this into Equation (1.13), and dividing both sides by $P_{i,t}$, yields the following expression for the average survival rate of cohort

$$s_{i,t} = \theta_{i,t} s^H_{i,t} + (1-\theta_{i,t})\, s^L_{i,t}. \tag{1.23}$$

Recall that $\theta_{i,t} = \frac{P^H_{i,t}}{P_{i,t}}$ is the share of high-skilled cohort $i$ at time $t$. This expression shows that the overall survival rate is a weighted average of the survival rates for high- and low-skilled individuals.

In the absence of data on high-skilled survival rates, the high-skilled survival rate is assumed analogously to Equation (1.17), using the following expression

$$s^H_{i,t} = (1+\psi_{i,t})\, s_{i,t}, \tag{1.24}$$

and low-skilled survival rate is

$$s^L_{i,t} = \frac{1-(1+\psi_{i,t})\,\theta_{i,t}}{1-\theta_{i,t}} s_{i,t}. \tag{1.25}$$

where $\psi_{i,t} \in \left[-1, \frac{1}{s_{i,t}} - 1\right]$ is an exogenous, time-dependent shock. This approach ensures that all survival rates, as well as the cohort-level weighted average, remain within the interval $[0, 1]$.

### 1.4.3 Physical capital accumulation

In the standard Solow framework, a fixed share of aggregate income is saved each period to finance capital accumulation. In contrast, the setting in this paper assumes that savings are derived from three sources: labor (disaggregated by skill type), land income, and return to physical capital. Each income component contributes to capital accumulation according to its own saving rate. As a result, the capital stock evolves over time according to the following law of motion

$$K_{t+5} = \sum_{i=15}^{65} \left(\sigma^H_i w^H_{i,t} P^H_{i,t} + \sigma^L_i w^L_{i,t} P^L_{i,t}\right) + \sigma^K r_t K_t + \sigma^X q_t X_t + (1-\delta) K_t, \tag{1.26}$$

where $\delta$ denotes the capital depreciation rate. $\sigma^K$ and $\sigma^X$ are the saving rates from capital and land income, respectively. $\sigma^H_i$ and $\sigma^L_i$ represent the saving rates of high-skilled and low-skilled workers in cohort $i$, respectively.

In the absence of direct data or literature on household saving behavior in Cambodia, the values for $\sigma^H_i$ and $\sigma^L_i$ are calibrated using cohort-level data from Figure A.2 in Deaton and Paxson (2000), which examines age-specific saving rates in Thailand in 1976. The resulting age profile of saving rates is taken as a reasonable approximation for the Cambodian context, based on the cultural and economic similarities between the two countries, particularly with regard to saving for essential expenditures such as healthcare, education, and family support.[17] Further discussion of

[17] It is important to note that Thailand implemented a universal healthcare system in 2002, which may have

these assumptions and data limitations is provided in the Chapter 1 Appendix.

Equation (1.26) an be rearranged using the expressions for factor incomes from Section 1.4.1.2, and simplified to yield

$$K_{t+5} = \widetilde{\sigma}_t K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta} + (1-\delta)K_t, \tag{1.27}$$

where

$$\widetilde{\sigma}_t = \beta\widetilde{\Omega}_t + \alpha\sigma^K + (1-\alpha-\beta)\sigma^X, \tag{1.28}$$

and

$$\widetilde{\Omega}_t = \frac{\sum_{i=15}^{65} \gamma_i \Phi_{i,t}^{\rho-\eta} \left[\sigma_i^H \lambda \theta_{i,t}^{\eta} + \sigma_i^L (1-\lambda)(1-\theta_{i,t})^{\eta}\right] P_{i,t}^{\rho}}{\sum_{i=15}^{65} \gamma_i \Phi_{i,t}^{\rho} P_{i,t}^{\rho}}. \tag{1.29}$$

The term $\widetilde{\sigma}_t$ denotes the effective saving rate from all factor incomes such as labor, capital, and land, and evolves over time with changes in income distribution and labor force composition. Unlike the fixed saving rate in the standard Solow model, $\widetilde{\sigma}_t$ is endogenous and varies across periods according to the structure of income sources and the demographic makeup of the workforce. The component $\widetilde{\Omega}_t$ reflects average savings out of labor income, weighted by cohort size, productivity, and skill composition. It increases when high-skilled, higher-saving individuals make up a larger share of the working population. By linking saving behavior to the underlying demographic and economic structure, this framework provides a more realistic and dynamic representation of capital accumulation.

## 1.5 Initial values and parameter calibration

This section presents the calibration strategy, including the parameter values and assumptions used to simulate the benchmark and counterfactual scenarios. Parameters are sourced from existing literature or chosen to ensure consistency with the model.

### 1.5.1 Initial capital stock

The model is initialized in 1950 under the assumption that the economy was already on a balanced growth path prior to that year. This assumption is necessary due to the lack of data for Cambodia before 1950. Initial conditions for capital, labor, and land are reflected this steady state, with each factor assumed to growth at constant rate, denoted by $g_K$, $g_L$, and $g_X$, respectively. To approximate the 1950 growth rates of labor and population, the model uses average growth rates observed between 1955 and 1970. From 1955 onward, growth rates evolve according to the dynamic equation. The growth rates of capital and land productivity are calculated to ensure consistency with the balanced growth assumption. Additionally, the steady state the growth rate of GDP per capita is set to 0.011. A full list of parameter values is provided in Table A.1.

subsequently reduced household reliance on private savings.

The initial capital stock at steady state is given by

$$K_0 = \left[ \left( \frac{\widetilde{\sigma}_0}{g_K + \delta} \right) L_0^{\beta} X_0^{1-\alpha-\beta} \right]^{\frac{1}{1-\alpha}}, \tag{1.30}$$

where $\widetilde{\sigma}_0$ is the initial effective saving rate. Details on the derivation of these values are provided in the Chapter 1 Appendix.

Assuming the economy was on a balanced growth path as of 1950 allows the model to focus on the demographic shock associated with the Khmer Rouge regime. This assumption is due to the absence of economic and demographic data for Cambodia prior to 1950. It also serves as a modeling simplification and should not be interpreted as a literal historical claim. The goal is to isolate the long-run effects of the genocide from broader transitional dynamics. Without this assumption, the model would also need to account for earlier historical transitions and convergence dynamics, which would complicate the model structure and make it harder to isolate the specific economic effects of the Khmer Rouge period.

### 1.5.2 Initial $\theta_{i,t}$

The steady-state share of high-skilled newborns depends on the probabilities that children born to high-skilled and low-skilled parents become high-skilled themselves. This is captured by the expression

$$\theta_0 = \frac{\pi^L}{1 - (\pi^H - \pi^L)}, \tag{1.31}$$

which plays a central role in determining consistent values for $\pi^H$ and $\pi^L$. At the steady state, $\theta_0$ is assigned to 12.42 percent based on the observed share of high-skilled individuals in the 2010 Cambodia Socio-Economic Survey (CSES) (see the Chapter 1 Appendix for derivation and calibration details). Holding $\theta_0$ constant, a high value of $\pi^H$ requires a correspondingly lower value of $\pi^L$, and vice versa, to maintain the same steady state skill composition.

For example, if $\pi^H = 0.8$, then $\pi^L$ must be very low. In this case, the majority of high-skilled children are born to high-skilled parents, meaning that the reproduction of skilled labor depends heavily on this group. After a mortality shock that disproportionately affects high-skilled individuals, the Khmer Rouge genocide, the limited contribution of low-skilled parents slows the recovery of $\theta_{i,t}$. The stock of future skilled workers becomes constrained with fewer high-skilled parents remaining, and the population takes longer to rebuild its skill base.

In contrast, if $\pi^H = 0.3$, $\pi^L$ must be relatively high to maintain the same value of $\theta_0$. This indicates that low-skilled parents play a much greater role in producing high-skilled children. In this case, even after a substantial loss of high-skilled population, the recovery of $\theta_{i,t}$ is faster because the remaining low-skilled population continues to contribute significantly to the development of new high-skilled individuals.

More generally, when $\pi^H < 0.5$, the model implies that high-skilled reproduction is more evenly distributed across the population, making the economy more responsive to shocks that disproportionately affect high-skilled adults. Conversely, when $\pi^H > 0.5$, the transmission of skill becomes more heavily reliant on high-skilled parents. As a result, shocks that disproportionately affect this group lead to deeper and longer-lasting disruptions in human capital formation, delaying the recovery of the high-skilled population (see Figure A.18, A.19, and A.20).

In the baseline calibration, the model sets $\pi^H = 0.3$ and $\pi^L = 0.1$, reflecting a more inclusive structure of skill transmission. This setup allows for a more robust and realistic recovery of $\theta_{i,t}$ following the 1975 mortality shock (see Table A.7 for economic outcomes under alternative values of $\pi^H$ and $\pi^L$).

### 1.5.3 Parameters

The factor shares in the production function are assigned standard values commonly used in the literature: $\alpha = 0.3$, $\beta = 0.6$, and $1 - \alpha - \beta = 0.1$, following Ashraf et al. (2013). Growth rates for labor, capital, and land are set at the inferred balanced growth rates as described above. Additional details on the calibration procedures and parameter choices are provided in the Chapter 1 Appendix.

The depreciation rate $\delta$ is set to 0.07, consistent with Karra et al. (2017), Schmitt-Grohé and Uribe (2005). Saving rates for capital $\sigma^K$ and land $\sigma^X$ are both set at 0.0855 and assumed constant across time, following Ashraf et al. (2013), Heston et al. (2012).

In addition, saving rates for high-skilled labor $\sigma_i^H$ and low-skilled labor $\sigma_i^L$ are calibrated with the assumption that high-skilled individuals save at twice the rate of low-skilled individuals. This is motivated by human capital theory and the tendency for higher-income individuals to save more in the absence of universal health care or social insurance. In Cambodia, as in Thailand, households rely heavily on personal savings to cover essential expenses, making it reasonable to expect higher saving rates among high-skilled individuals.

Moreover, to make the model more realistic, all saving rates are assumed to be zero during the genocide period. This reflects the historical reality in which individuals were forced to work for minimal rice rations, with no access to wages or monetary income, and thus no capacity to save. More generally, the model follows the structure of the Solow growth framework, where saving rates are treated as exogenous parameters, not derived from utility maximization or intergenerational decision-making. This approach improves transparency and tractability, particularly when simulating periods in which individuals could not engage in normal economic activities and market systems were no longer functioning (see Table A.3).

Furthermore, the weighted parameter, $\gamma_i$, are assumed to growth at a constant and are normalized to sum to one across working ages (see Table A.3). These weights reflect a concave age-earnings profile, increasing with experience and declining later in life.

The elasticity of substitution between labor from different age cohorts $\rho$ is set to 0.9 to approximate near-perfect substitution in aggregate effective labor. Similarly, the elasticity of substitution between high-skilled and low-skilled labor within each age group $\eta$ is set to 0.9. These assigned values reflect the assumption that skill types are imperfect substitutes.

The relative productivity of high-skilled labor compared to low-skilled labor $\lambda$ is set to $\frac{2}{3}$. It is calibrated using relative wages of high-skilled and low-skilled workers, normalized to the wage of low-skilled individuals in cohort 15, based on data from the 2010 CSES. Also, $\lambda$ is adjusted to ensure that the average wage predicted by the production function matches the average wage observed in the data.

### 1.5.4 Population dynamics parameters

To generate the simulated population time path for both the actual and counterfactual scenarios, age-specific migration factors are calculated as residuals using Equation (A.1.17) due to the lack of migration data disaggregated by age cohort. Aggregate net migration in a given year is defined as $\sum_{i=0}^{95}(m_{i,t}-1)P_{i,t}$.

For the counterfactual scenario, age-specific migration factors are assumed to equal the average values observed over the 1950-2015 period, rather than being interpolated across missing years. This approach is motivated by the fact that, in the absence of the genocide, the true pattern of migration is unknown and difficult to infer credibly. Using historical averages provides a neutral and stable benchmark, avoiding the need to impose speculative assumptions about age-specific migration flows. It also helps to eliminate the sharp emigration observed during the genocide and the surge in return migration and refugee resettlement that occurred in the post-genocide period, particularly during the 1990s (see Chapter 1 Appendix for calibration procedures details).

A one-time shock parameter $\psi_{i,t}$ is set to $-0.25$ only for cohort $i \in \{15, 20, ..., 95\}$ in the year 1975, and zero otherwise. This shock represents a 25% mortality penalty for high-skilled individuals, based on historical evidence that educated, affluent, and urban populations were disproportionately targeted by the Khmer Rouge regime (Clayton, 1998; Holck & Cates Jr, 1982; Kiernan, 1996). Since approximately 25% of the urban population perished during this period, applying this shock to high-skilled cohorts is a reasonable approximation.

While the model assumes high mortality for all adults during the genocide, it imposes an additional penalty on high-skilled individuals to reflect their disproportionate vulnerability. The survival rate for low-skilled cohorts is then adjusted so that the overall survival rate aligns with observed data (see Equation 1.23). This does not imply that low-skilled adults were unaffected, but rather that they faced relatively lower risk of being targeted. This assumption alters the post-genocide skill composition by disproportionately reducing the share of surviving high-skilled individuals.

In the absence of the genocide, survival rates for years 1970, 1975, and 1980 are treated as missing in order to remove the mortality shock from the simulation. This adjustment is applied equally to both high-skilled and low-skilled survival rates. The missing values are then interpolated using a cubic spline method based on survival rates from adjacent years. This technique ensures smooth and demographically plausible survival paths over time.

The same approach is applied to age-specific fertility rates, with values from 1970 to 1990 treated as missing. The missing values are then interpolated using a cubic spline method to construct a smoothed counterfactual fertility series. This adjustment removes the effects of both the genocide-related fertility disruption and the post-genocide replacement fertility observed during that period. By doing so, the model isolates long-run demographic patterns that would have prevailed in the absence of the conflict, providing a more stable baseline for counterfactual analysis.

## 1.6 Result

This section presents the simulation results, comparing the actual and counterfactual demographic and economic outcomes. The analysis focuses on population structure, skill composition, and output per capita over time.

### 1.6.1 Population simulation results: Actual vs counterfactual

This section compares the actual demographic path of Cambodia to a counterfactual scenario in which the genocide did not occur.

Figure 1.4: Total population time path.

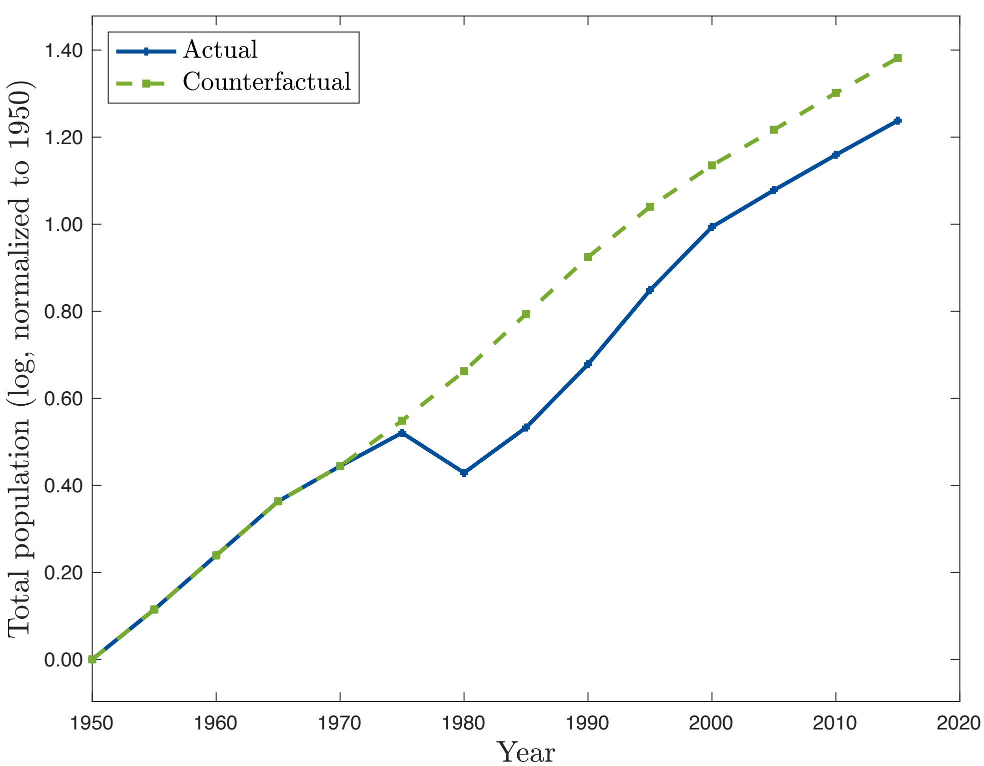
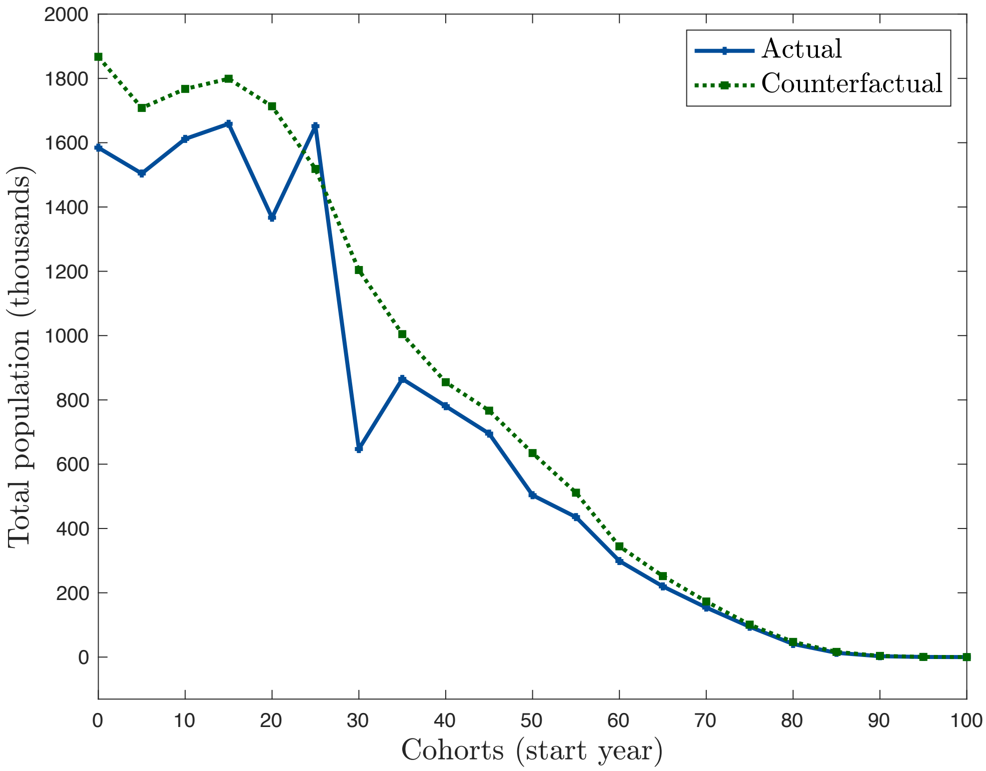



Figure 1.4 presents the time paths of total population under the actual and counterfactual scenarios. The two series begin to diverge around 1975, reflecting the historical disruptions in the actual case. By 1980, the difference between the two paths reflects excess mortality caused by the Khmer Rouge, estimated at approximately 1.8 million. This estimate is consistent with earlier findings in Heuveline (1998, 2015), Kiernan (1996). If one accepts the age distribution assumptions in this model, this number can be interpreted as the excess deaths caused by the genocide. Indeed,

the demographic effects of the genocide persist in the long run, even though population growth rates converge in the post-crisis period.

Figure 1.5 compares the actual and counterfactual age distributions in 2010. The counterfactual distribution is much smoother and less the volatility seen in the actual data, especially among cohorts aged 15 to 30.

Figure 1.5: Age distribution in 2010.




The difference in the cohort aged 30 to 34 reflects the effects of post-genocide replacement fertility, while the smaller size of the 35 to 39 age group points to a severe fertility decline during the Pol Pot period. Banister and Johnson (1993) estimate that there were approximately $216,000$ fewer births in 1979 due to the genocide. Meanwhile, the shape of the distribution for ages 0 to 14 is almost identical between the two scenarios, suggesting that the longer the time horizon, the more the effects of the genocide dissipate (see Table A.4).

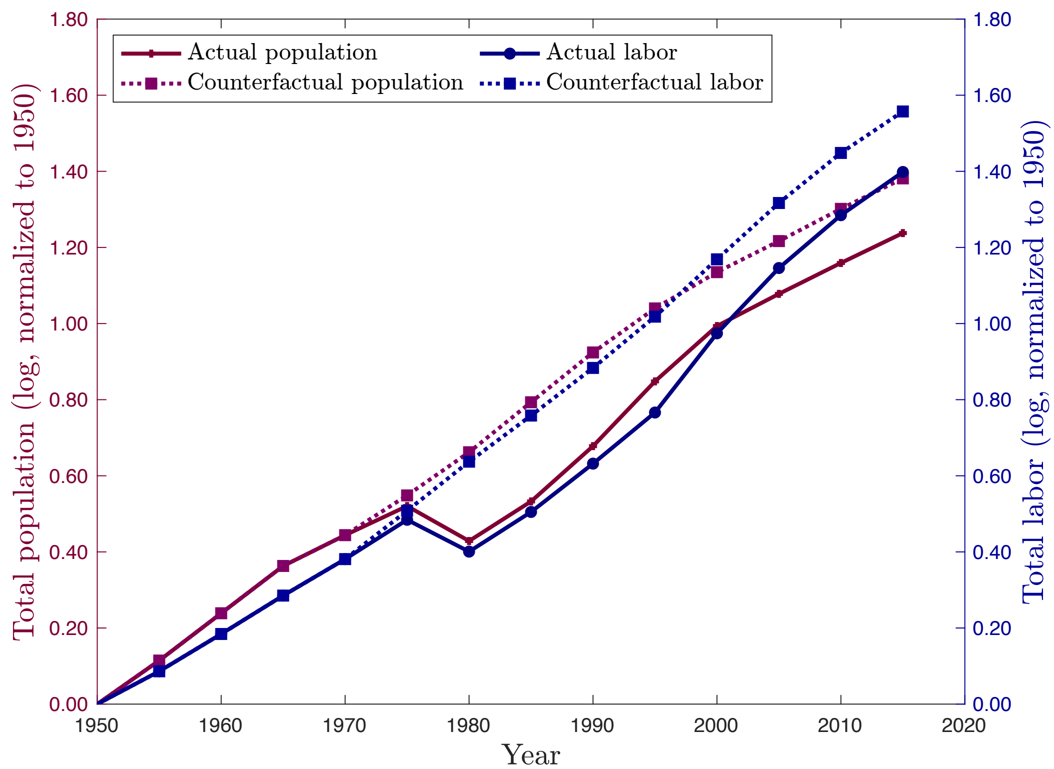
Together, the reconstructed counterfactual paths offer a smooth picture of how Cambodia's standard demographic might have evolved under more stable conditions, allowing for a meaningful comparison with actual post-genocide outcomes. These demographic differences form the foundation for analyzing long run effects on output per capita, labor composition, and capital accumulation in the following sections.

#### 1.6.1.1 Population by skill type

The disproportionate targeting of the educated population during the Pol Pot regime led to a sharp divergence in survival rates by skill level. Figure A.3 shows that, in the actual scenario, survival rates for high skilled individuals aged 15 and older were significantly lower than those of the low skilled population around 1975, reflecting the one-time mortality shock associated with the genocide. In contrast, the counterfactual survival rates for both skill groups exhibit smooth patterns consistent with the overall counterfactual survival trend, since they are constructed without the genocide-related shock (see the dash line in Figure A.15).

Figure 1.6 presents the total population and skill-heterogeneity working-age population time paths. After 2000, the working age population begins to grow more rapidly than the total population, despite replacement-level fertility and declining mortality. This divergence indicates a demographic recovery, as the working-to-non-working age ratio gradually converges toward the counterfactual path (see Figure 1.8).

Figure 1.6: Total population and labor time path.



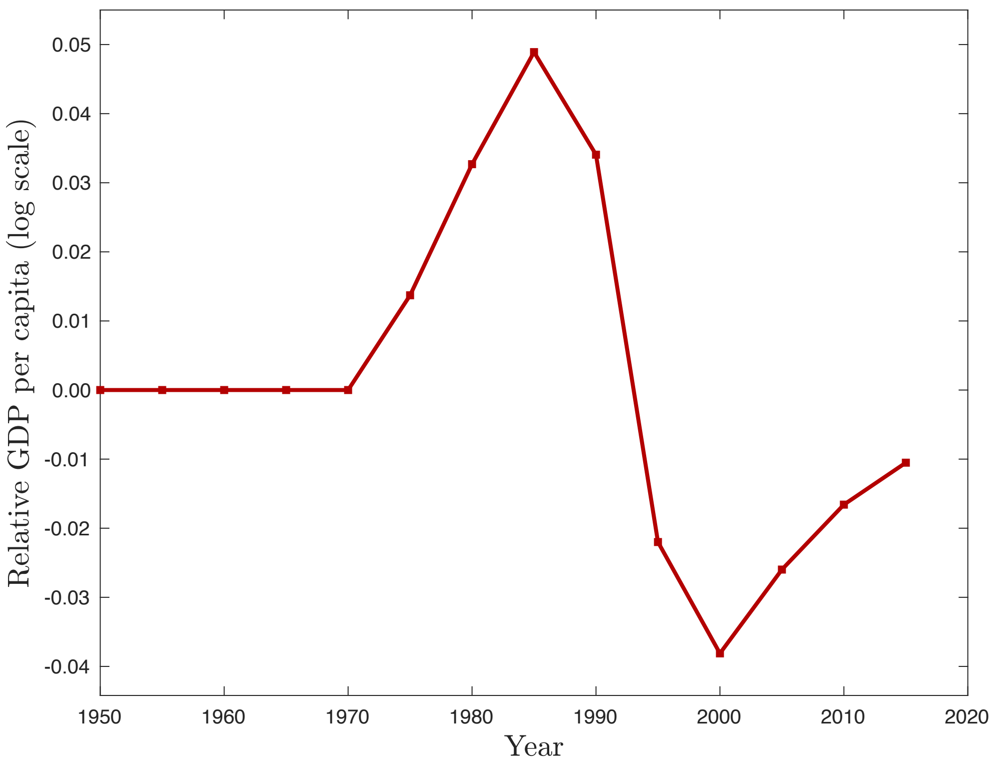

Although the genocide caused lasting damage to the age and skill structure of the population, the data suggest that the labor force has slowly begun to recover. The growing share of the working age population, particularly among lower skilled cohorts, signals a post-crisis stabilization phase. However, gaps in cohort size and skill composition remain, with long run implications for human capital accumulation and labor market productivity.

### 1.6.2 Baseline result

This section presents the baseline simulation results from a model that assumes a closed economy with no technological progress. Saving behavior is held constant over time for capital and land income, while labor-saving rates vary by cohort and skill. The results compare actual demographic and economic outcomes with a counterfactual scenario in which the Cambodian genocide did not occur.

Figure 1.7 shows time path of relative GDP per capita from 1950 to 2015.[18] Between 1950 and 1970, actual and counterfactual GDP per capita are nearly identical, reflecting the fact that both economies share the same population dynamics and savings behavior in the absence of the genocide.

Figure 1.7: Relative GDP per capita.




In 1975, actual GDP per capita exceeds the counterfactual (see Figure 1.7). This is largely explained by a temporary capital-labor advantage due to population loss. Excess mortality likely affected the young and elderly more than the working-age population, temporarily increasing the share of individuals participating in the labor force. As a result, the actual working-to-non-working age ratio is higher than in the counterfactual scenario.

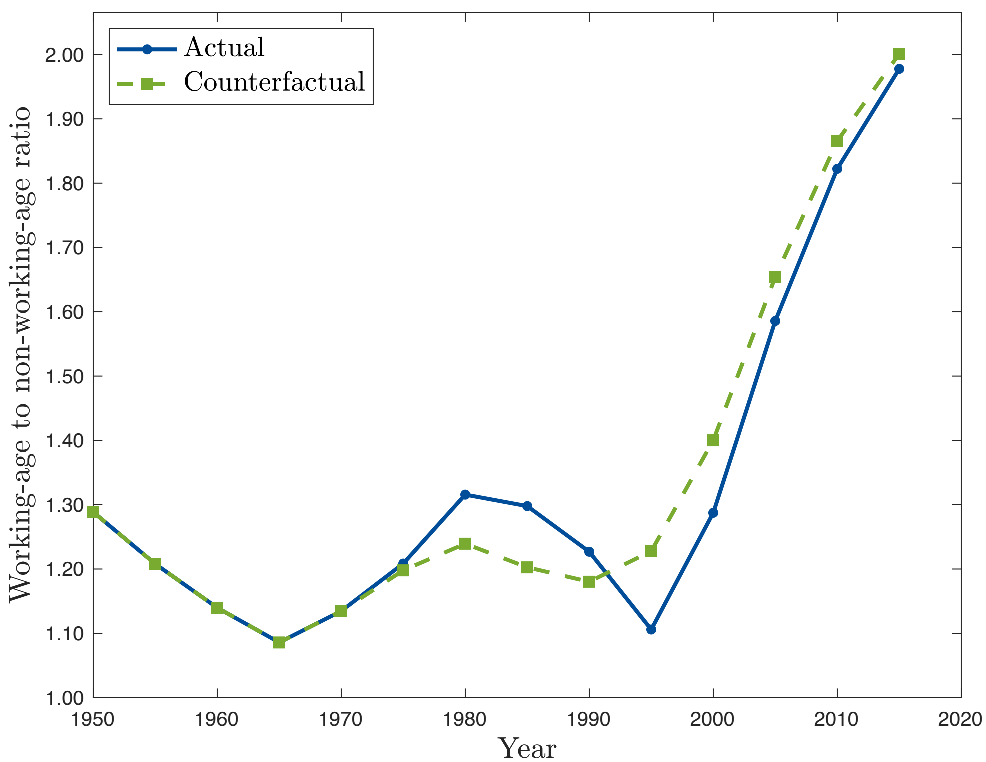
Between 1980 and 1985, the gap persists, driven by higher land-to-labor and capital-to-labor ratios in the actual economy. Fewer workers imply more land and capital per person, yielding short-run productivity gains (see Figure A.22 and A.21). In addition, a higher saving rate contributes

[18] The relative GDP per capita is calculated as the actual GDP per capita divided by the counterfactual GDP per capita at each point in time.

further to the temporary increase in relative GDP per capita. These patterns demonstrate that capital, being more flexible than labor in the short run, plays a stronger role in the immediate post-genocide recovery.

In 1995, the counterfactual GDP per capita becomes higher than the actual. This shift is driven by the ongoing demographic imbalance, especially in the labor-to-population ratio (see Figure A.23). Short-run fertility replacement increases the number of dependents, lowering the actual working-to-non-working age ratio compared to the counterfactual (see Figure 1.8). Although the baby boom cohort enters the labor market, it does not fully replace the pre-genocide working-age population. In contrast to physical capital, which can be rebuilt over a few periods, population recovery, especially in terms of age structure and skill composition, takes much longer. This highlights the importance of demographic balance. A larger non-working-age population can reduce output, even when capital accumulation rebounds. These results are consistent with prior studies such as Ashraf et al. (2013), Bloom and Canning (2008), Bloom et al. (2011), which show that economies tend to grow faster when the working-age share of the population is higher.

Figure 1.8: Ratio of working-age to non-working-age population.




After 1995, the actual economy starts to catch up. The working-to-non-working age ratio gradually improves, indicating a slow demographic recovery. However, the long-run effects of the genocide persist. The loss of high-skilled parents continues to reduce the stock of skilled labor in future generations through intergenerational channels. Table 1.1 shows that in 2010, both GDP and GDP per capita remain higher in the counterfactual scenario about 15% and 1.6%, respectively. Although the actual economy continues to recover, convergence with the counterfactual may occur as long-term demographic distortions are reduced over time. Note that Table A.7 also presents results under alternative values of $\eta$ and $\rho$.

Table 1.1: Relative GDP and GDP per capita (2010): Baseline model.

| Substitution parameters | Relative GDP | Relative GDP per capita |
|---|---|---|
| $\rho = 0.9$ and $\eta = 0.9$ | 0.8532 | 0.9836 |
| $\rho = 0.9$ and $\eta = 0.5$ | 0.8434 | 0.9723 |
| $\rho = 0.5$ and $\eta = 0.9$ | 0.8493 | 0.9791 |
| $\rho = 0.5$ and $\eta = 0.5$ | 0.8401 | 0.9685 |

*Notes:* Relative GDP measures the total output in 2010 of actual GDP relative to counterfactual scenarios, with relative GDP per capita defined analogously. The production function uses factor shares of $\alpha = 0.3$ for capital, $\beta = 0.6$ for labor, and the remaining share $1-\alpha-\beta = 0.1$ for land. During the Pol Pot regime (1975–1979) in the actual scenario, saving rates from all three income sources (capital, labor, and land) are set to zero.

These findings are similar to the demographic dividend literature. As shown by Ashraf et al. (2013), Bloom and Canning (2008), Bloom et al. (2011), the economic outcomes tend to improve when fertility and mortality decline together, increasing the share of the working-age population. In addition, when the sizes of young and older adult cohorts are more evenly balanced, productivity growth are more likely to occur (Feyrer, 2007). Although Cambodia's standard of living has gradually improved since the fall of the Khmer Rouge, the genocide continues to have a lasting negative impact on long-term economic outcomes. This results is in line with the findings from Acemoglu et al. (2011) on the economic legacy of the Holocaust in Russia. However, this study differs by explicitly comparing actual outcomes to a counterfactual scenario in which genocide did not occur.

#### 1.6.2.1 Special case I: Without skill heterogeneity

To isolate the contribution of skill composition to long-run economic outcomes, this special case sets the productivity parameter $\lambda = 0.5$, treating high-skilled and low-skilled labor as equally productive. As a result, the model assumes a homogeneous labor force and removes the distinction between skill types. The production function remains the same as in Equation (1.1), but labor input depends on cohort size and age structure. Equation (1.4) simplifies to $P_{i,t}$, and the skill adjustment factor become $\Phi_{i,t} = 1$. Thus, Equation (1.5) becomes

$$L_t = \left[ \sum_{i=0}^{95} \gamma_i P_{i,t}^{\rho} \right]^{\frac{1}{\rho}}. \tag{1.32}$$

All other parameters are held constant, following the baseline specification in Section Baseline result (see Table A.1 for all parameter values). Additional derivations and calibration details are provided in the Chapter 1 Appendix.

By removing human capital differences, this setting highlights how demographic structure alone, without skill heterogeneity, affects post-genocide economic outcomes. Although both the baseline model and Special Case I show a temporary increase in actual GDP per capita relative to the counterfactual after the genocide, the underlying mechanisms differ. In Special Case I, short run gains are driven by increases in land-to-labor and labor-to-population ratios, as the model does not account for the productivity losses associated with the destruction of the educated population.

Figure A.26 shows the log of actual GDP per capita relative to the counterfactual. The pattern is overall similar to the baseline. However, in 1980, actual GDP per capita exceeds the counterfactual, as the capital-per-worker effect no longer plays a role. In this Section, the land-to-labor and labor-to-population effects dominate. In 1985, the saving effect becomes more significant and leads to a sharp increase in actual GDP per capita relative to the counterfactual.

In 1990, however, the saving effect begins to weaken, and actual per capita income starts to decline, despite positive contributions from land-labor, capital-labor, and labor-population ratios. In 1995, actual income per capita falls below the counterfactual level, primarily due to the negative effect of the genocide on the labor-to-population ratio. After 1995, the actual economy begins to catch up as the baby boom cohort enters the labor force. Table 1.2 shows that actual GDP per capita in 2010 remains approximately 0.5 percent below the counterfactual. This gap reflects how the labor-to-population effect outweighs gains from capital and land reallocation in the actual scenario. Note that Table 1.2 also includes outcomes for $\rho = 0.5$.

Table 1.2: Relative GDP and GDP per capita (2010) without skill heterogeneity: Special case I.

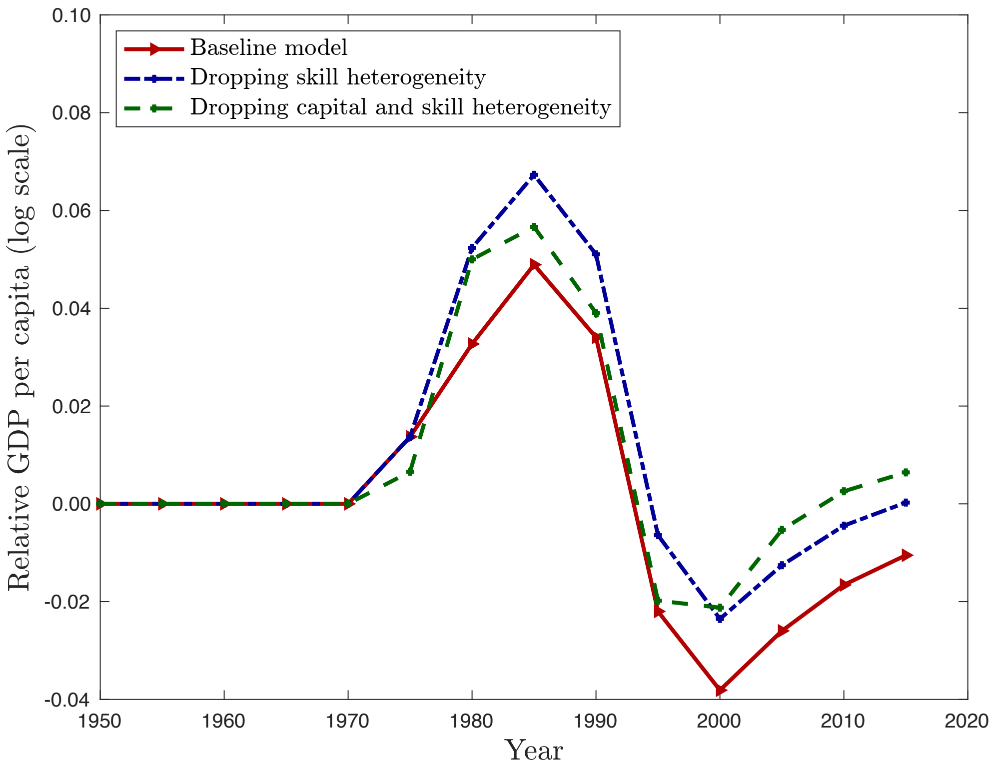
| Substitution parameters | Relative GDP | Relative GDP per capita |
|---|---|---|
| $\rho = 0.9$ | 0.8636 | 0.9956 |
| $\rho = 0.5$ | 0.8593 | 0.9905 |

*Notes:* Relative GDP measures the total output in 2010 of actual GDP relative to counterfactual scenarios, with relative GDP per capita defined analogously. The production function uses factor shares of $\alpha = 0.3$ for capital, $\beta = 0.6$ for labor, and the remaining share $1-\alpha-\beta = 0.1$ for land. During the Pol Pot regime (1975–1979) in the actual scenario, saving rates from all three income sources (capital, labor, and land) are set to zero.

Figure 1.9 compares the long-run relative GDP per capita under the baseline model with heterogeneous labor and the special case with homogeneous labor. In 2010, the blue dashed line (homogeneous labor) lies above the red solid line (heterogeneous labor). It indicates a faster recovery when skill composition is excluded. The gap between the two lines reflects the skill composition effect. Also, this gap is driven by the loss of educated individuals during the genocide and the slow intergenerational transmission of human capital after the genocide. In the baseline model, these losses lead to persistently lower labor productivity and slower output growth. In contrast, the homogeneous labor model does not capture the consequences of losing skilled workers, and therefore it may underestimate the long-term economic cost of the genocide and overstate the speed of recovery.

These findings suggest that even in the absence of skill heterogeneity, demographic shocks can significantly alter growth trajectories through changes in capital and labor compositions. However, incorporating skill heterogeneity provides a more accurate account of the long-lasting damage caused by the loss of human capital, which continues to suppress productivity and output growth over time.

Figure 1.9: Relative GDP per capita across heterogeneity scenarios.




#### 1.6.2.2 Special case II: Without skill heterogeneity and capital

In this setting, both capital and skill heterogeneity are removed from the model. The productivity parameter is set to $\lambda = 0.5$, and the capital share $\alpha = 0$, meaning that the capital share is fully transferred to labor. The resulting production function includes only labor and land inputs, following a structure similar to the Malthusian economy. Under these conditions, Equation (1.1) simplifies to

$$Y_t \quad = \quad L_t^{\beta} X_t^{1-\beta}. \tag{1.33}$$

The initial $X_t$ is normalized to 1 similar to the baseline model. The share of land in output remains 0.1. Parameter values are consistent with the baseline calibration unless otherwise noted (see Table A.1 and Chapter 1 Appendix).

In this simplified framework, the impact of demographic changes on output is determined directly by the age distribution, as suggested by Denton and Spencer (1973). Special Case II provides a useful benchmark for comparing actual and counterfactual paths of GDP and GDP per capita based purely on population structure and the allocation of labor and land, and it is independent of skill or capital accumulation.

Figure A.27 shows actual GDP per capita relative to the counterfactual from 1950 to 2015. Between 1950 and 1970, both economies follow a balanced growth path. From 1975 to 1990, actual

GDP per capita exceeds the counterfactual, driven by a higher working-age to non-working-age ratio in the actual economy. This shift is the result of a morality shock affecting children and the elderly during the Khmer Rouge period. At the same time, excess deaths increase the land-to-labor ratio, and thus the land effect becomes more dominant in driving output per worker. This dynamic is consistent with a Malthusian framework, where a smaller population temporarily raises income per capita due to greater land availability per worker.

From 1995 onward, the counterfactual economy begins to exceed the actual economy in output, consistent with the trend observed in the baseline model. Table 1.3 shows that in 2010, counterfactual GDP is about 13 percent higher than actual GDP. However, counterfactual GDP per capita is slightly lower by approximately 0.3 percent. This result suggests that living standards in the actual economy are marginally higher. This outcome reflects the fact that the positive effect of greater land availability per person outweighs the negative impact of a lower labor-to-population ratio. In the absence of capital, land becomes the primary driver of income per capita. A smaller population therefore results in more land per person, which contributes to higher per capita income and improved living standards (see Figure 1.4). Note that Table 1.3 also includes outcomes for $\rho = 0.5$.

Table 1.3: Relative GDP and GDP per capita (2010) without capital and skill heterogeneity: Special case II.

| Substitution parameters | Relative GDP | Relative GDP per capita |
|---|---|---|
| $\rho = 0.9$ | 0.8697 | 1.0026 |
| $\rho = 0.5$ | 0.8643 | 0.9963 |

*Notes:* Relative GDP measures the total output in 2010 of actual GDP relative to counterfactual scenarios, with relative GDP per capita defined analogously. The production function uses factor shares of $\beta = 0.9$ for labor, and the remaining share $1-\beta = 0.1$ for land.

Figure 1.9 shows the long-run relative GDP per capita across all three model specifications. The gap between the red solid line (baseline) and the green dashed line (Special Case II) reflects the combined effects of capital accumulation and skill composition. The gap between the green and blue dashed lines (Special Case I) isolates the effect of capital. As a result, Special Case II cannot capture the productivity losses from human capital destruction or the delayed recovery driven by capital-skill complementarity.

Although skill heterogeneity is excluded in this special case, comparisons with the baseline model help illustrates the importance of skill composition for long-run growth. Figure 1.9 shows that lost of educated individuals under the Pol Pot regime continue to alter economic outcomes. These effects are intergenerational, as human capital takes longer to rebuild than physical capital. Table 1.1 shows that the economic cost of the genocide, measured through the skill composition channel, is more severe when cohorts are less substitutable or when the elasticity of substitution between high-skilled and low-skilled labor is lower. These results demonstrate the long-term consequences of demographic shocks, particularly when they disrupt the stock of human capital.

Together, the baseline model and the two special cases reveal the relative importance of different production inputs and demographic mechanisms. The baseline model incorporates both capital and skill heterogeneity, and it captures the full interaction between physical and human capital. Special Case I removes skill differences to isolate the effects of demographic structure and capital intensity. Special Case II simplifies the model further by excluding both capital and skill composition, and focuses on the role of labor and land allocation. Comparing outcomes across the three settings shows that short-run differences are largely driven by changes in age structure and land availability. However, long-run divergence is explained by the persistent effects of lost human capital and reduced intergenerational skill transmission.

## 1.7 Conclusion

This paper provides a quantitative analysis of how the Cambodian genocide impacts the country's demographic structure and long-term economic growth. By building a counterfactual scenario in which the genocide did not occur, the analysis isolates the effects of excess mortality, fertility decline, and changes in age structure and skill composition. The model integrates these demographic inputs into a production function with land, capital, and skill-heterogeneous labor, capturing both the short-term dynamics and long-term growth outcomes.

The baseline results show that the Cambodian genocide had both short-run and long-run effects on economic outcomes. In the immediate aftermath of the genocide, actual GDP per capita temporarily exceeds the counterfactual due to a demographic shock that increases the share of working-age individuals and raises land and capital per worker. These factors contribute to short-term gains in productivity. However, this effect is largely mechanical, resulting from a smaller population and a temporarily higher working-age-to-non-working-age population ratio. This short-term demographic dividend should not be mistaken for genuine economic recovery, as it risks understating the true economic cost of genocide. Beginning in the mid-1990s, the counterfactual economy overtakes the actual, as it benefits from a more balanced demographic structure and skill composition. The long-run divergence is primarily driven by the persistent loss of skilled cohorts and the slow recovery of intergenerational skill transmission, which continue to suppress productivity and limit output growth in the actual economy.

Special Case I and II allow for a deeper decomposition of the mechanisms driving growth differences. When skill differences are removed, as in Section 1.6.2.1, the capital-labor and land-labor channels remain effective. However, the wage inequality channel disappears due to the absence of productivity variation across skill types. When capital is excluded, as in Section 1.6.2.2, the model approximates a Malthusian economy in which income is primarily determined by land per capita and demographic structure. Together, these variations show that short-run outcomes are largely influenced by resource allocation and age composition. In contrast, long-run divergence is driven by the persistent effects of human capital loss and the low elasticity of substitution between skill groups.

These results highlight how demographic shocks can have persistent effects on long-run economic recovery. While the model shows that GDP per capita exceeds the counterfactual path shortly after the genocide, this effect is largely mechanical. It is driven by a smaller population and a temporarily higher working-age-to-non-working-age ratio. This short-term demographic dividend, however, should not be mistaken for genuine economic recovery and understating the true economic cost of genocide.

This study shows that the effects of genocide extend far beyond immediate population loss. They persist through demographic imbalances, reduced human capital, and lower labor productivity, which together contribute to a lower level of economic development. The findings contribute to the literature on demographic shocks and the long term impacts of war and conflict, aligning with studies on the demographic dividend (Ashraf et al., 2013; Bloom & Canning, 2008) and the economic legacy of conflict (Acemoglu et al., 2011). By constructing a counterfactual scenario, the analysis isolates structural mechanisms such as skill composition and labor substitutability, and provides a benchmark for understanding what Cambodia's economic trajectory might have looked like without the genocide. While the analysis offers valuable insight into the long-term demographic and economic consequences of the Cambodian genocide, the results should be interpreted with caution. The model assumes full labor force participation and omits other structural changes, such as international trade dynamics or institutional reforms. These results are not intended to imply causal claims or specific policy recommendations, but rather to illustrate what the Cambodian economy might have looked like had it not experienced the genocide.

# Chapter 2

# The Effects of Being Born in Urban Areas During the Cambodian Genocide

## 2.1 Introduction

Malnutrition in infancy and related early-life shocks can have profound and lasting effects on human capital accumulation and long-term socioeconomic outcomes. A growing body of theoretical and empirical studies, including the influential work of and Cunha and Heckman (2007), Cunha et al. (2006), and Heckman et al. (2006), emphasizes that early childhood is a sensitive period for the development of both cognitive and noncognitive skills. These abilities evolve through self-productivity, where early skills enhance the development of later ones, and dynamic complementarity, in which early investments increase the productivity of subsequent investments. Within this framework, a negative health shock early in life can trigger a cumulative process of disadvantage. These disadvantages often manifest as lower educational attainment, poorer health, and reduced labor market outcomes.

Several empirical studies supports these mechanisms. For instance, Meng and Qian (2009) find that exposure to China's Great Famine significantly reduced adult height and education attainment. Similarly, Alderman et al. (2006) link early-life drought and conflict in Zimbabwe to lower educational attainment. Barham (2012) shows that health interventions in Bangladesh improve cognitive skills and schooling. Yi et al. (2015) find that early health shocks lead to poorer academic performance among affected twins through both biological and household investment channels. In the Indian context, Attanasio et al. (2020) demonstrate that early health deficits significantly constrain cognitive development. While parental investments enhance cognition at all ages, their effects are particularly strong during early childhood.

One of the most devastating episodes of forced displacement and institutional collapse in modern history occurred under the Khmer Rouge regime in Cambodia between 1975 and 1979, commonly referred to as the Cambodian Genocide. Led by Pol Pot, the regime forcibly relocated approximately two million residents, primarily from the capital city of Phnom Penh, to rural areas (Kiernan, 2002).[1] The educated, middle-class, and professional groups were explicitly targeted, resulting in the deaths of roughly 25% of the urban Khmer population. Urban dwellers tend to have better education, health, and higher socioeconomic status than the rural populations Holck and Cates Jr (1982).[2] As a result, the regime's so-called "zero urban population" policy had particularly severe effects on infants and young children, whose growth and cognitive development are especially sensitive to nutritional and environmental shocks during early childhood.

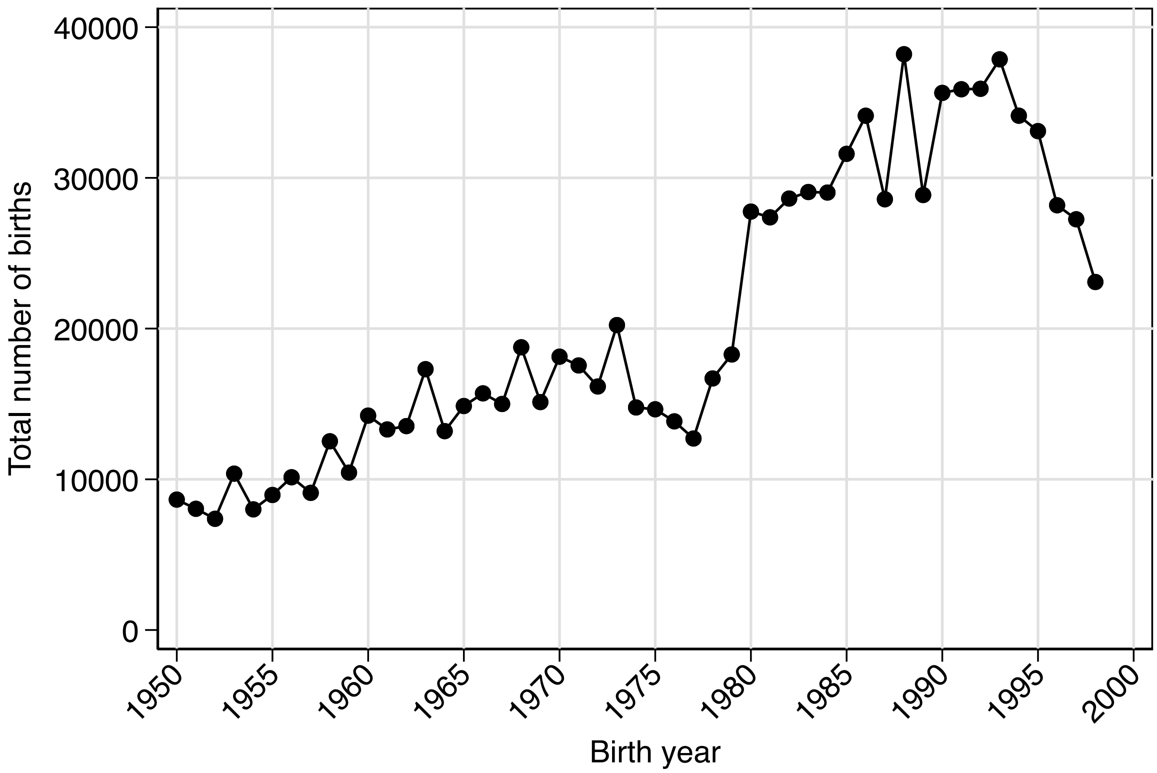
Individuals born in Cambodia during the Khmer Rouge years were likely exposed to severe food shortages, poor sanitation, and widespread health risks. One indication is that a large part of the overall population died from starvation. According to Ebihara and Kiernan (1993), food scarcity became increasingly severe after 1976-1977 for "New people", while the Khmer Rouge cadres and "Old people" had access to abundant food supplies.[3] Meng-Try (1981) reports that approximately $30,000$ young children died from malnutrition and poor sanitation in the last six months of 1975 alone. Most likely, newly born and in-utero children were particularly vulnerable. Birth cohorts from those years tend to be smaller than for neighboring years, suggesting that many newborns died and/or that many pregnancies ended in miscarriage. A significant decline in birth rates is evident across Cambodia (see Figure 2.1) and in all urban areas (see Figure B.1).

[1] According to Kiernan (2002) calculation, there were two million urban Khmer population in 1975, and half a million died. The death toll excluded other ethnic groups and rural Khmer.

[2] Holck and Cates Jr (1982) study the fertility and population of Khmer refugees in two Kapuchean refugee camps.

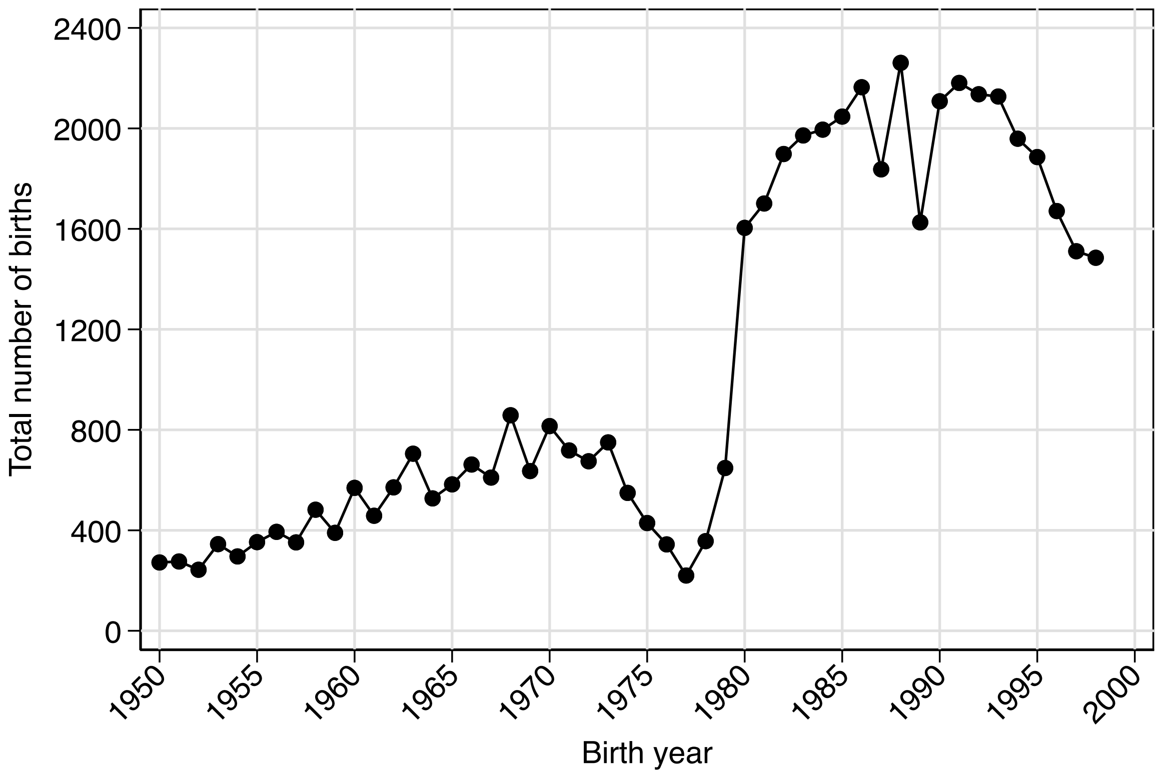
[3] The term "New people" (also referred to as 17th People) refers to individuals who were forcibly relocated to rural villages during the mass evacuation. In contrast, Old People (or Base People) were those who had already resided in the countryside prior to April 17, 1975, and were allowed to remain there.

Figure 2.1: Total number of births in Cambodia by birth year.




Moreover, this shock had a disproportional impact on urban populations, as the educated and professional classes were primary targets of the Khmer Rouge. This is also evident in census data, which show a much sharper decline in births for urban areas than rural ones. In Phnom Penh, birth rates fell significantly, with the 1977 cohort experiencing the steepest drop (see Figure 2.2).

Figure 2.2: Total number of births in Phnom Penh by birth year.




One challenge for this research is identifying a source of exogenous variation in malnutrition or related early-life shocks. This paper considers a relatively sudden nation-wide event, the Khmer Rouge regime in Cambodia (1975-1979), which likely caused undernutrition, particularly among urban populations. I focus on the long-term effects of being born in urban areas during this period, specifically on cohorts born between 1975 and 1979. To that end, I use individual-level census data, primarily from 1998, with additional analysis using other years. These data include information on respondents' birth year and district, as well as their educational attainment at the time of the census. I also construct a household wealth index, following the approach of Davila et al. (2014, 2022), using responses to questions on asset ownership, utilities, and dwelling characteristics.

I then use a two-way fixed effects regression to assess if whether individuals born in urban districts during the Khmer Rouge years have different educational and/or wealth outcomes later in life compared to those born in other parts of Cambodia and in other years. I consider different time periods to define the Khmer Rouge period, and I use a few different measures of how urban districts are (see Section 2.3.2). I also control for various potentially confounding factors. One limitation is that some individuals born to forcibly relocated urban families, known as "New people" under the Khmer Rouge, may be recorded in the data as rural-born because their parents had already been displaced at the time of birth. This misclassification may bias the estimate of the true urban effect. I discuss this issue further in the empirical section.

The main results are overall robust. Being born in an urban district during the Khmer Rouge reign has a negative effect on both wealth and education. Interpreting the Khmer Rouge event as a shock to nutrition and care, the results thus seem consistent with the effects described by Cunha et al. (2006).

As mentioned, to undertake this exercise, I need to assess if the respondents were born in an urban or rural district. To that end, I consider a few different approaches. The first and simplest one is to use an indicator for districts close to the capital and largest city in Cambodia, Phnom Penh. Second, I use the three largest cities, Battambang, Phnom Penh, and Siem Reap. A third approach is to use indicators for 31 districts that considered as urban today according to my main source from General Population Census of Cambodia 1998.

These are all binary measures, taking values of zero or one, and thus cannot capture variation in the degree to which different districts were affected by the Khmer Rouge regime. Moreover, how urban a district was before the rise of the Khmer Rouge may not be well proxied by how urban it is today. Part of the contribution of this paper is therefore to construct a couple of different measures of urbanization, based on how deep the dip in fertility rates was in each district around the Khmer Rouge era. Specifically, I measure the fraction of all births taking place in the district during the Khmer Rouge reign, relative to those happening in the same district over a longer time span before and after the event. Since I want to measure urbanization, I use one minus the ratio constructed. A deeper drop in fertility likely indicates a more urbanized district that was heavily disrupted during the regime.

As expected, these alternative measures show a positive correlation with my other (binary) indicators. The empirical results are also broadly similar when using these alternative measures in the regressions. They become particularly strong when controlling for district-specific cubic time trends, which allow the nonbinary measures to capture the sharpness of the dip in births during the Khmer Rouge period.

This paper contributes to a couple of different fields of literature. Most obviously, it related to the growing literature on the long-term impacts of political conflict on human capital. Several studies document that exposure to war or upheaval during childhood leads to persistent educational and economic losses. De Walque (2006) documents that the Khmer Rouge era devastated Cambodia's school system, resulting in lower educational attainment for cohorts of school age during 1975–79. Merrouche (2011) finds that landmine contamination in Cambodia led to reduced schooling but no significant impact on earnings. Islam et al. (2016, 2017) further demonstrate that the Khmer Rouge regime caused declines in both educational attainment and earnings, including intergenerational effects. Outside Cambodia, Akresh and de Walque (2008) find that the Rwandan genocide reduced schooling and primary completion rates. Similarly, Ichino and Winter-Ebmer (2004), Kesternich et al. (2014) show that childhood exposure to WWII significantly lowered education, earnings and long-term health outcomes. Meng and Gregory (2002) report that China's Cultural Revolution caused substantial educational losses among urban students due to school closures.

This paper is closely related to the work of Islam et al. (2016, 2017), who also examine the long-term impacts of the Khmer Rouge regime. Islam et al. (2016) study how exposure to civil conflict during primary school years (ages 9-14) affected individuals' education, earnings, and fertility later in life, using regional mortality rates as a proxy for conflict intensity. Islam et al. (2017) extends this analysis by investigating intergenerational effects, examining how parental exposure to the genocide influenced children's education, health, and marriage patterns. In contrast, my study focuses on a different cohort and research question. I examine individuals who were born in urban districts during the Khmer Rouge period (1975 to 1979), and assess how early-life exposure to forced displacement and deprivation affected their adult educational attainment and household wealth. Rather than relying on regional mortality as a proxy for exposure, I use the year and location of birth, specifically whether an individual was born in an urban district during the genocide years, as the treatment variable. This cohort-based approach offers a complementary perspective on how early-life shocks during conflict can produce long-lasting socioeconomic consequences.

The paper also relates to the broader literature linking early-life health shocks and malnutrition to long-run outcomes. Grantham-McGregor et al. (2007) show that children under five in developing countries are at risk of not reaching their developmental potential due to poverty, malnutrition, and stunting. Glewwe and Jacoby (1995) find that childhood malnutrition significantly delays school enrollment in Ghana. In Great Britain, Currie and Hyson (1999) show that low birth weight reduces educational attainment, as well as poorer self-reported health and labor market outcomes. To strengthen causal inference, several studies use natural experiments. Maccini and Yang (2009) exploit rainfall variation in Indonesia, finding women born in wet season attain better health, education, and socioeconomic outcomes. Akresh et al. (2012) use war exposure in Nigeria to show adolescent shocks reduce adult stature. Huang et al. (2013) extend these findings to older age, showing that both adult height and height shrinkage in China are linked to cognitive decline and social economic status in later life. Moreover, early age health is significantly influences cognitive performance and later-life earnings. Bhalotra et al. (2022) show that Sweden's 1930s infant health program improved children's cognitive skills and increased women's adult earnings by 19.5%, illustrating how early health investments can yield long-term returns in education and economic outcomes. Similarly, Chen et al. (2022) find that early exposure to tap water in rural China significantly improved children's cognitive test scores.

The structure of the paper is organized as follows. Section 2.2 describes the data source. Section 2.3 presents the empirical setting. 2.4 shows the empirical results. 2.5 offers conclusions.

## 2.2 Data

This paper uses the Integrated Public Use Microdata Series - International (IPUMS-I) census years 1998 as it is the first official census conducted after the end of Pol Pot era. The data represent 10 percent of the total population and are nationally representative.

The data is richer in information on birth years and birthplace districts. Of the 183 birthplace districts, 181 are included in the data, with the remaining two classified as danger zones after the downfall of the Khmer Rouge regime, so the census enumeration could not be conducted. This information is based on the General Population Census of Cambodia 1998 Final Census Results from the National Institute of Statistics, Ministry of Planning.[4]

Since the IPUMS-I does not classify birthplace districts into urban or rural, the urban district of birthplaces is classified based on information obtained from General Population Census of Cambodia 1998. Consequently, a total of 31 districts of the birthplaces are identified as urban (see Table B.28). Furthermore, Figure 2.4 provides a visual representation of the district locations, including urban, rural, and big three urban cities. Note that the big three urban cities are Battambang, Phnom Penh, and Siem Reap, and these cities are selected based on their higher urban population density.

IPUMS-I dataset provides detailed information at the household and individual levels, but it does not have data on income and expenditures. Nonetheless, the data includes information regarding owners' asset characteristics, enabling the construction of wealth index scores.

These scores serve as a proxy for household socioeconomic status. They are consistent with wealth indices constructed using Demographic and Health Survey (DHS) data, as demonstrated by Davila et al. (2014) and Davila et al. (2022). According to the authors, it is necessary to have at least 30 asset ownership variables to measure the wealth index score better. Hence, the 2008 Census is used for validity check as it has more variables than the 1998 census.

## 2.3 Empirical methodology

In this section, the paper outlines the empirical approach utilized to examine the impact of cohort effects on the outcomes of years of schooling and wealth index scores. The paper focuses only on the cohorts born between 1965-1984.

### 2.3.1 Empirical model

This empirical specification adopts a generalized Difference-in-Differences (DiD) framework, exploiting variation in exposure across birth cohorts and geographic location to estimate the causal impact of being born in an urban area during the Khmer Rouge regime. It compares individuals born during the regime in urban districts (the treatment group) to those born either in rural areas or in other time periods (the control groups). By interacting the temporal and spatial dimensions of exposure, the approach isolates the long-term effects of early-life displacement, institutional collapse, and political violence.

To estimate the effects of being born in an urban area during the Khmer Rouge period, the following regression is estimated

$$Y_{ibd} \quad = \quad \alpha + \beta_1 C_b U_d + \beta_2 X_{ibd} + \gamma_d + \lambda_b + \sum_{k=1}^{3} \delta_k^d b^k + \varepsilon_{ibd}, \tag{2.1}$$

[4]The General Population Census of Cambodia 1998 can be retrieved from http://www.nis.gov.kh/index.php/km/15-gpc/87-general-population-census1998

where $Y_{ibd}$ denotes the outcome of interest either the years of schooling or wealth index score for individual or household $i$, respectively, born in birth year $b$ in birth district $d$, and $b \in [1965, 1984]$.

The interaction term $C_b U_d$ is the main variable of interest, capturing the causal effect of being born in urban area during the Pol Pot years. $C_b$ is a cohort effect dummy, defined using alternative specification births during 1975-1979 or 1976-1978, or only 1977. $U_d$ denotes either urban district dummies or a continuous urbanization measure. The coefficient $\beta_1$ estimates the differential impact on educational attainment or wealth in adulthood for for urban-born cohorts during the genocide relative to other cohorts and rural-born individuals.

$X_{ibd}$ is a vector of controls including the marital status, number of children, and sex. Sex is an indicator of female and male otherwise. Marital status includes categories for married, divorced, or widowed, with single as the reference group.

District fixed effects $\gamma_d$ control for time-invariant characteristics at the district level, such as geographic conditions, cultural norms, or historical infrastructure, and remove cross-district unobserved heterogeneity. Birth year fixed effects $\lambda_b$ capture national-level cohort shocks or events affecting all individuals born in a given year.

$\sum_{k=1}^{3} \delta_k^d b^k$ denotes district-specific cubic time trends, which allow each district to follow a smooth, nonlinear trends over time. These trends account for unobserved district-level variation in outcomes, including factors such as demographic shifts, institutional developments, and post-war recovery processes, which typically evolve gradually and vary across districts. This flexibility enables the regression to account for how each district changed over time, helping to mitigate potential confounding from endogenous district-level dynamics.

As shown in Figure B.3, the evolution of average years of schooling by birth cohort exhibits a non-monotonic pattern across districts. Thy show sharp drops during the Khmer Rouge period 1975-1979, followed by uneven and nonlinear recovery trends. This visual evidence supports the use of cubic, rather than linear or quadratic, time trends to flexibly capture district-specific temporal variation. These trends absorb nonlinear dynamics that could otherwise bias the estimated effect of urban birth during the genocide.

The specification addresses omitted variable bias by incorporating district and birth-year fixed effects and by allowing for separate, smooth cubic trends in each district. Standard errors are clustered at both the district and cohort levels to allow for within-cluster correlation. While birth-year fixed effects control for national-level cohort shocks, they do not capture local district-level dynamics that evolve differently across space. District-specific cubic trends are essential in this post-conflict context, where recovery varied significantly across regions. These trends help isolate the treatment effect from other endogenous developments occurring at the district level.

It is important to note that, due to the mass evacuation of cities in 1975, the vast majority of urban-origin families were forcibly relocated to rural areas. Urban districts were largely cleared of civilians, and only regime cadres, military personnel, and administrative staff were permitted to remain. However, a memoir by one of the Khmer Rouge's senior leaders suggests that a limited number of technically skilled individuals may have been selectively returned to urban areas. According to Khieu Samphan, during the first year of Central Committee meetings, proposals were made to recall technicians to operate factories in Phnom Penh and to allow some intellectuals to return for technical education projects, such as the creation of a vocational training school (Khieu, 2004, p.58). This historical evidence supports the interpretation that individuals born in urban districts during the Khmer Rouge period were likely children of regime insiders or individuals deemed valuable to the regime, rather than displaced urban-origin families. This distinction has important implications for interpreting the treatment variable, which captures exposure to urban areas at birth rather than family background. If cadre families experienced better conditions during the regime, this may bias the estimated treatment effect upward and lead to an underestimation of the true disadvantage experienced by displaced urban-origin children born in rural areas.

Moreover, a potential identification concern arises from the Khmer Rouge's policy of forcibly relocating urban residents to rural areas beginning in 1975. Some of these displaced individuals gave birth during the regime, and their children would appear in the dataset as having been born in rural districts. This misclassification may reduce the observed difference between treatment and control groups, thereby potentially biasing the estimated treatment effect toward zero. As a result, the reported coefficients could understate the true impact of being born in an urban area during the Khmer Rouge period. Children of urban-origin families are recorded as rural-born, even though they likely faced similar or greater deprivation associated with their parents' urban status. This systematic misclassification may lead to an underestimation of the true negative effect of urban born during this period.

Although detailed information on parents' district of residence prior to the Pol Pot regime is unavailable, this study defines the urban indicator based on the district of birth, capturing the individual's geographic exposure at birth rather than parental origin. While this does not fully eliminate concerns related to forced displacement, it ensures internal consistency by defining treatment status according to location rather than family background. This approach aligns with the study's aim to estimate how being born in an urban area, rather than to urban-origin parents, affected long-term outcomes during a period of institutional collapse. Also, the fixed effects and district-specific cubic time trends, helps account for differences in local recovery patterns and other unobserved temporal dynamics that may correlate with displacement.

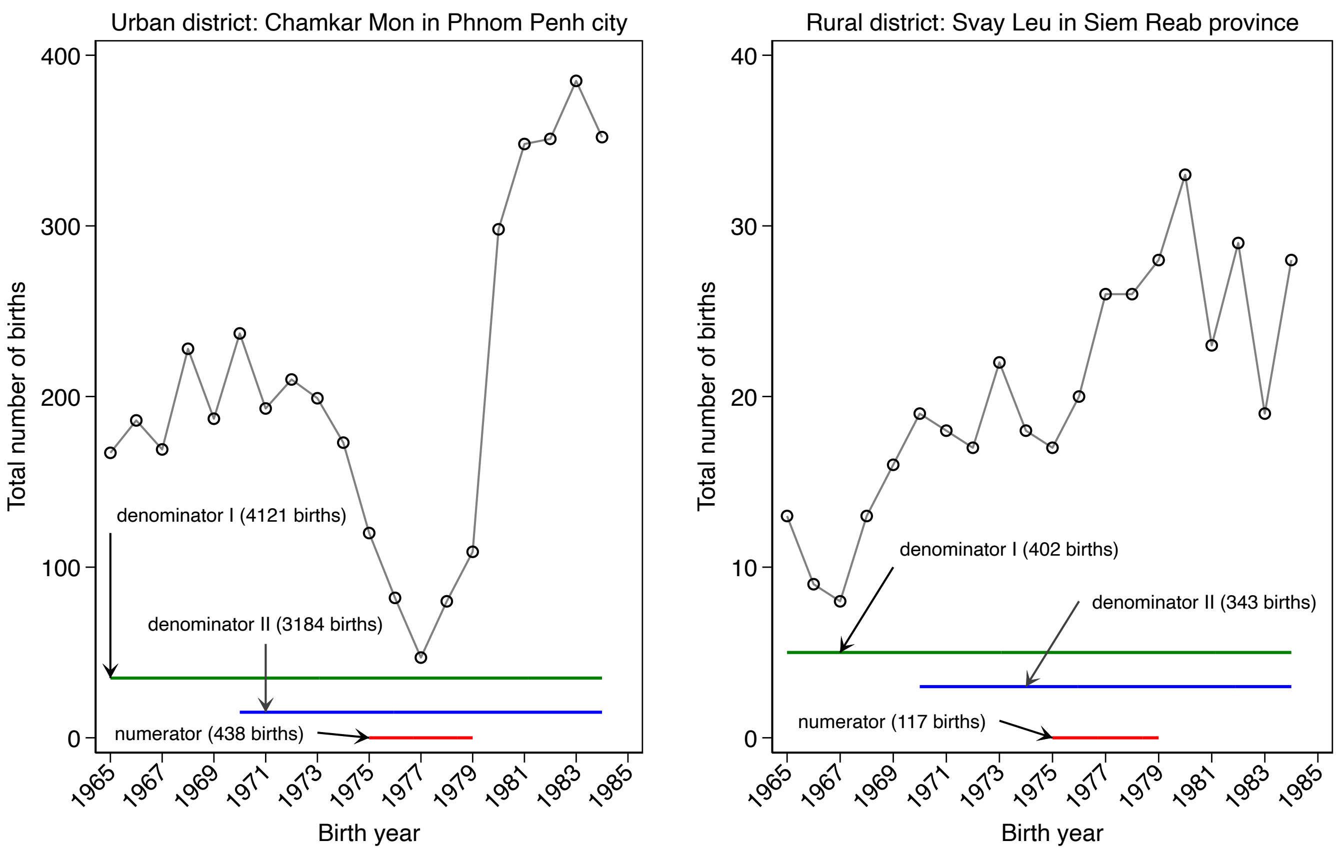
Taken together, although forced relocation introduces a potential source of measurement error in the treatment variable, this bias would likely attenuate the estimated effect. Therefore, the results presented in this study should be interpreted as conservative estimates of the true negative impact of urban-born during the genocide, rather than as inflated or spurious findings.

### 2.3.2 Measure of urbanization

Urbanization $U_d$ is measured using two distinct methods. The first constructs continuous urbanization proxies based on the relative dip in birth rates during the Khmer Rouge period. The second method defines urban status using binary urban district indicators derived from official administrative classifications.

The first method measures the magnitude of the decline in births during the Pol Pot years within each district. The underlying assumption is that areas with a larger birth dip were more urbanized, as the regime's forced evacuation policy disproportionately affected urban populations. There are two variants of this proxy, urban proxy I and urban proxy II (see Figure 2.3). The distinction between urban proxies I and II is their denominators as urban proxy I uses the entire interval of birth years from 1965-1984 (see the green line), while urban proxy II restricts the interval to 1970-1984 (see the blue line).

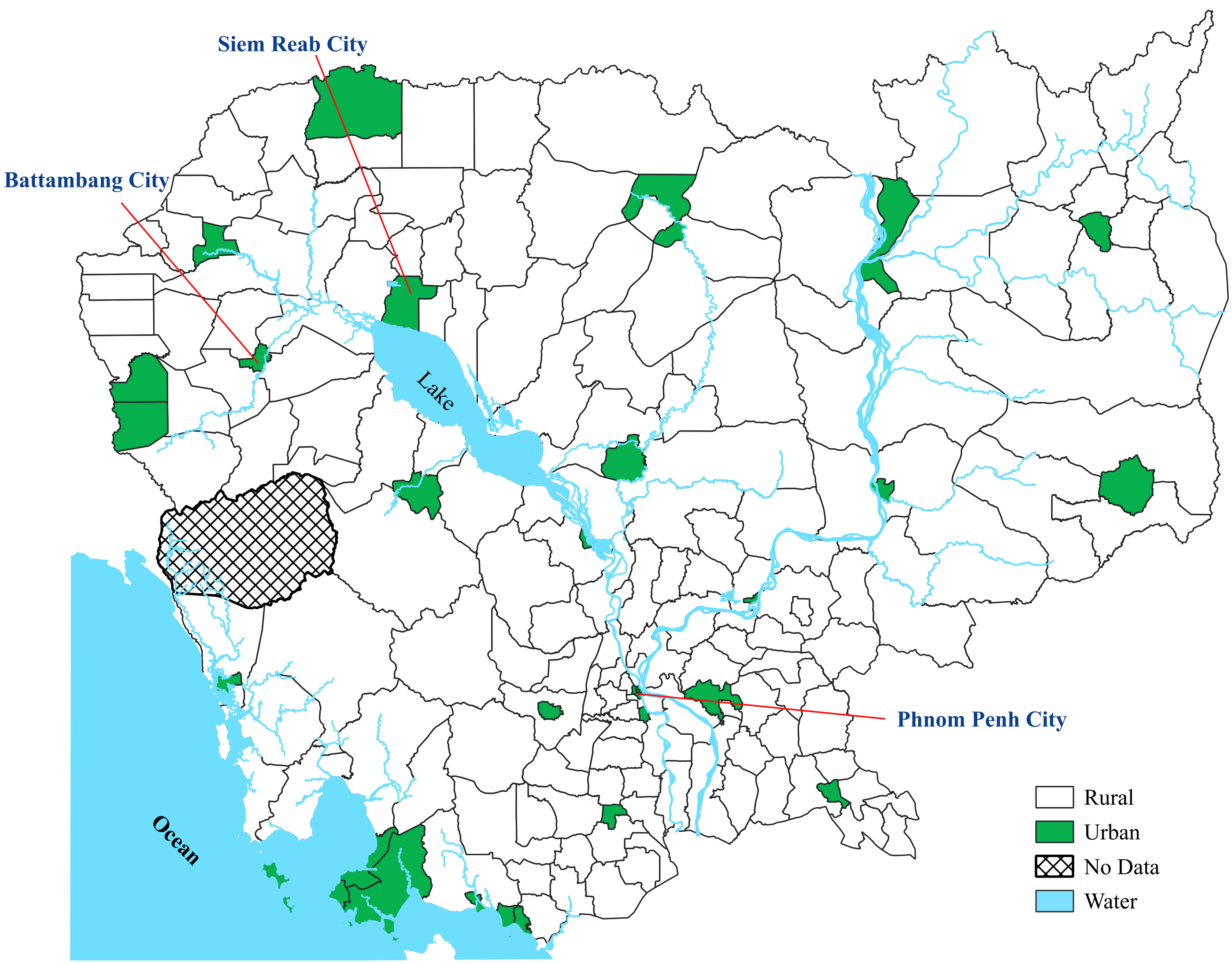
Figure 2.3: Urban proxy $\mathrm{I}_{75-79}$ vs urban proxy $\mathrm{II}_{75\text{-}79}$.




Note: urban proxy $\mathrm{I}_{75\text{-}79}$ equals 1 minus the length of red line (numerator) divided by the length of the green line (denominator I). urban proxy $\mathrm{II}_{75\text{-}79}$ equals to 1 minus the length of the red line (numerator) divided by the length of the blue line (denominator II).

Urban proxy I is defined as $\left(1 - \frac{\text{Born}_b}{\text{TotBorn}_{65\text{-}84}}\right) \times 100$, where $\text{TotBorn}_{65\text{-}84}$ is total number of birth between 1965-1984. Urban proxy II uses a narrows baseline and is defined as $\left(1 - \frac{\text{Born}_b}{\text{TotBorn}_{70\text{-}84}}\right) \times 100$, where $\text{TotBorn}_{70\text{-}84}$ is then total number of birth between 1970-1984. In both expressions, $\text{Born}_b$ represents the total umber of births in a specific year or short interval $b$, typically 1975-1979, 1976-1978, or 1977 alone. A smaller birth ratio, $\frac{\text{Born}_b}{\text{TotBorn}}$, reflects a larger relative dip in births. Subtracting this ratio from one yields a higher measurement of urbanization. Thus, districts that experienced larger birth declines during the regime are inferred to have been more urbanized prior to the Khmer Rouge period.

Figure 2.4: Location of urban district of birthplaces.




Note: In total, 31 out of 183 districts are classified as urban. The definition of urban districts is based on information from the National Institute of Statistics, Ministry of Planning, as reported in the General Population Census of Cambodia 1998: Final Census Results.

For example, the left panel in Figure 2.3 illustrates Chamkarmon, a typical urban district in Phnom Penh. In contrast, the right panel shows Svay Leu, a rural district in Siem Reap province. The urbanization measurement is applied similarly in both panels. However, the left panel displays a significant dip in the total number of births during the Pol Pot years, particularly from 1975 to 1979, with a more pronounced decline in 1977. By contrast, the right panel shows a peak in births in 1977. This increase is attributed to the forced relocation of urban populations into rural areas under the regime's zero-urban-population policy. Consequently, the left panel indicates higher urbanization than the right, as evidenced by its smaller birth ratio.

The second method uses urban districts as dummies. The first dummy contains 31 urban districts, including Battambang, Phnom Penh, and Siem Reap cities (see Figure 2.4 for location of the urban and rural birth district). The second dummy focuses more narrowly on the six urban districts located within Cambodia's three largest urban centers, namely Battambang, Siem Reap, and Phnom Penh. These cities are selected based on their higher urban population density.[5] The third dummy isolates Phnom Penh alone, the country's capital and most heavily urbanized city. Among its seven administrative districts, four are classified as urban and are used in this paper.

Table B.1 in the Chapter 2 Appendix reports the summary statistic of main variables used in this paper. Table B.2 reports the correlation between average years of schooling, wealth index scores, urban proxies I and II, and the urban dummy variables. Both educational attainment and household wealth are positively and significantly correlated with all measures of urbanization, with significance at the 1% level. Furthermore, Figure B.4 of the Chapter 2 Appendix illustrates the correlation between the continuous urban proxies and the urban dummies, revealing a strong positive relationship. This alignment suggests that the two methods for measuring urban status are likely capturing a similar underlying urbanization pattern.

### 2.3.3 Constructing wealth index score

In the absence of income and expenditure data, this study constructs a household wealth index as a proxy for socioeconomic status. The index is derived using the first principal component analysis (PCA), based on household asset ownership, dwelling characteristics, and access to utilities (see Table B.29 for the lists of characteristics). This approach follows the methodology of Davila et al. (2014, 2022), who demonstrate that PCA-based indices using census data are valid and consistent with wealth measures constructed from Demographic and Health Survey (DHS) data.

In particular, the wealth index based on asset based follows general form

$$WI_{it} = \sum_{j=1}^{n} \kappa_{jt} w_{ijt}, \tag{2.2}$$

where $WI_{it}$ is the wealth index predicted for household $i$ in census year $t$. Note that the unit observation is the household, not the individual because the asset ownership attaches to the head of the household. $w_{jit}$ stands for asset $j$ for household $i$ in time $t$. $\kappa_j$ is the weight assigned to asset $j$ derived from the first PCA. The first PCA captures the maximum variance compared to other components. Note that $j \in [1, n]$ where $n$ equals 22 in 1998, (see Table B.29 in Chapter 2 Appendix).

Table B.3 reports the summarized statistic of wealth index scores, and its scores are standardized normalization. Out of the $212,967$ observations, only $67,605$ are used in the regression analysis

[5] According to the Population 1998 census, Krong Preah Sihanouk, Krong Kaeb, and Krong Pailin are also classified as entirely urban. However, they are excluded from this analysis because of their smaller population size and limited urban extent during the period of interest. Instead, this analysis focuses on the most demographically dense cities.

because the paper, again, focuses on the birth years between 1965 and 1984. Although the wealth index scores serve as the proxy for socioeconomic status, they may not accurately represent actual wealth or socioeconomic status. This limitation arises from the fact that the asset ownership information in the data contains only information about utility availability, dwelling characteristics, the number of rooms, or the number of household members, as mentioned by Davila et al. (2014, 2022). However, Table B.2 shows a robust and statistically significant positive relationship between wealth index scores and urban proxies I and II. Thus, the wealth index score is more potential to capture the income and expenditure outcomes of the cohort urban born during the genocide.

## 2.4 Empirical results

This section presents the results of estimated schooling years and wealth index scores through a simple statistical framework.

### 2.4.1 Result of estimating years of schooling

Tables 2.1 presents results where years of schooling is regressed on an indicator for being born in 1977 (Cohort dummy 77) interacted across panels A-E with different measures of how urban the birth district is. The different columns vary the set of controls, with column (6) allowing for district-specific cubic time trends. All columns include a full set of fixed effects for both birth cohorts and birth districts. The estimate in Panel A and column (1), using the continuous urban proxy I to interact with the cohort effect, the estimated coefficient is statistically significant at the 1% level.

Columns (2) - (4) incorporate female, marital status, and the number of children, respectively. The estimated coefficients remain highly statistically significant and show similar magnitudes. Furthermore, column (5) includes all the control variables in columns (2) - (4). Despite a change in value from $-0.20$ to $-0.18$ compared to column (1), the estimated coefficient remains robust and statistically significant. Finally, column (6) additionally accounts for district-specific cubic time trends. These trends are appropriate given the non-monotonic evolution of schooling across birth years, as shown in Figure B.3. Hence, the estimated coefficient becomes larger and highly statistically significant.

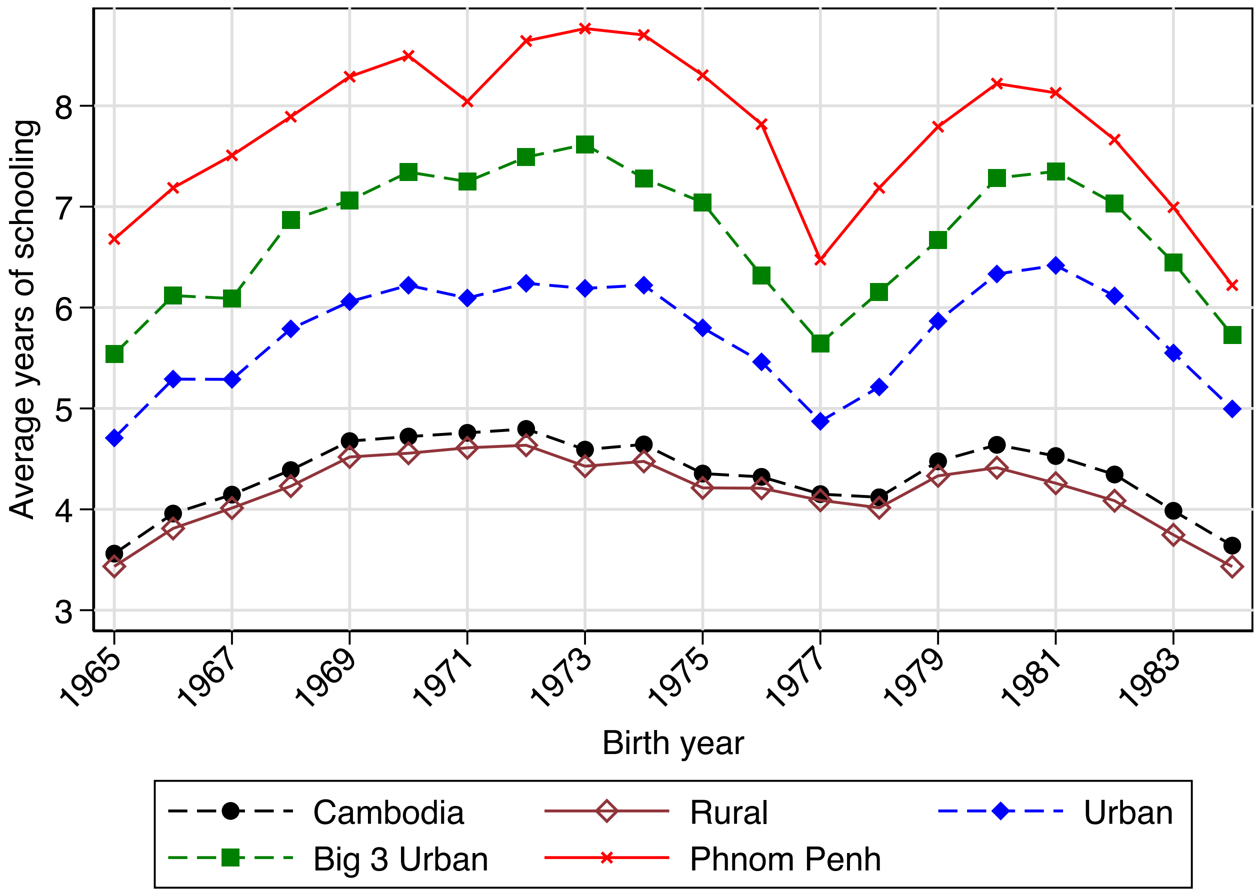
Panel B uses the interaction term between cohort 1977 dummy and urban proxy II variable. Although the estimated coefficients are slightly smaller than those in panel A, which is not surprising given its bigger ratio by the construction of the relative size of the dip (see Section 2.3.1), they are statistically significant for all columns. Moving on to panel C, the interaction between urban and cohort 1977 dummy is used. In all columns, the estimated coefficients are still negatively correlated with schooling and are larger than those in panels A and B. It suggests that individuals born in urban districts in the cohort 1977 have an average schooling attainment of approximately 0.40-0.60 lower than others. Despite the estimations, panels A, B, and C reflect the results shown in the blue line in Figure 2.5, and they consistently indicate that the more urban the birthplace, the less educational attainment. Yet, it is lesser for the cohort born in 1977.

Table 2.1: Years of schooling and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.20*** | -0.19*** | -0.18*** | -0.19*** | -0.18*** | -0.23*** |
| | (0.044) | (0.044) | (0.042) | (0.041) | (0.041) | (0.036) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel B | | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.16*** | -0.15*** | -0.15*** | -0.15*** | -0.14*** | -0.19*** |
| | (0.035) | (0.034) | (0.033) | (0.032) | (0.032) | (0.030) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel C | | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.50*** | -0.46*** | -0.50*** | -0.46*** | -0.43*** | -0.57*** |
| | (0.062) | (0.066) | (0.065) | (0.070) | (0.074) | (0.088) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel D | | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.56*** | -0.52*** | -0.56*** | -0.50*** | -0.47*** | -0.74*** |
| | (0.104) | (0.104) | (0.104) | (0.109) | (0.111) | (0.176) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel E | | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -1.27*** | -1.23*** | -1.29*** | -1.19*** | -1.18*** | -1.63*** |
| | (0.103) | (0.098) | (0.090) | (0.092) | (0.089) | (0.093) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square (OLS) regressions is estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual from the 1998 Census survey. The total number of observations is 384,044. The urban proxy I$_{77}$ is constructed as $(1-\frac{PopBorn_{77}}{Totpop_{65-84}})\times 100$. The urban proxy II$_{77}$ is defined as $(1-\frac{PopBorn_{77}}{Totpop_{70-84}})\times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban districts within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for the illustration.
$^*$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Figure 2.5: Average years of schooling by birth year.




In panel D, the cohort effect interacts with the big three urban cities. The estimates suggest that being born in the big three cities, the cohort of 1977 born experiences lower schooling of between 0.5 to 0.7. On the other hand, panel E, uses the Phnom Penh dummy to interact with the cohort 1977 and shows that the educational attainment is severely lower for the cohort effect than others. The educational attainment loss for this cohort ranges from approximately 1.3 in column (1) to about 1.6 in column (6), with highly statistically significant. The estimated coefficient in column (6) suggests that someone born in 1977 in one of the Phnom Penh districts had 1.63 fewer years of schooling than those born into other cohorts and/or other districts. This can be compared to the mean years of schooling in the full sample of around 6.5 years (see Figure 2.5). Additionally, the results from panels D and E correspond to the patterns represented by the green and red lines in Figure 2.5, respectively. The larger estimated coefficients in absolute terms in panel E can be attributed to the deeper dip in schooling observed for those born in 1977 in Phnom Penh, as the red line shown in Figure 2.5. The deeper dip is likely explained by the active involvement of their parents in the conflict and their parents' educational background because, during the regime, only Pol Pot leaders and soldiers were permitted to reside in the city, many of whom were less educated peasants.

Moreover, Figure B.5 in the Chapter 2 Appendix displays the estimated coefficients derived from regressing years of schooling on the individual's interaction with urban proxies or dummies. These regressions correspond to column (6) in Table 2.1 but encompass individuals born between 1965 and 1984. The figure reveals a consistent disadvantage associated with being born during the Pol Pot years and in urban districts, particularly the 1977 cohort born. Previous studies have demonstrated the profound impact on schooling and earnings for cohorts exposed to war and conflict during their primary and elementary school years (for Cambodian genocide, see De Walque (2006), Islam et al. (2016), and Merrouche (2011); for World War II, see Ichino and Winter-Ebmer (2004); for Chinese cultural Revolution, see Meng and Gregory (2002)). However, the cohort born in urban districts during the Khmer Rouge regime appears to endure the most adverse effect on educational attainment.

Furthermore, Tables B.4 and B.5 follow a similar structure to Table 2.1, with Table B.4 using the cohort 1976-1978 dummy and Table B.5 using the cohort 1975-1979 dummy. All estimated coefficients in column (6) exhibit high statistical significance and larger magnitudes when district-specific cubic time trends are included. These results consistently point toward the adverse effects of being born in urban areas during the genocide, irrespective of statistical significance. However, individuals born in Phnom Penh between 1976-1978 and 1975-1979 experience significantly lower years of schooling than other cohorts.

Further evidence separately obtained by examining female and male urban-born individuals confirms the negative impact on years of schooling. The findings consistently demonstrate that the higher the urbanization level of the birthplace, the lower the educational attainment for the cohort born during the Pol Pot era, particularly the cohort born in 1977 (see Tables B.8, B.9, B.10, B.11, B.12, and B.13). These results are consistent with the adverse effects observed in Tables 2.1, B.4, and B.5, regardless of statistical significance. Additionally, using the 2008 census data, the cohort born in urban areas during the historical episode, especially those born in 1977, encounters significantly lower levels of educational attainment. Notably, the lowest educational attainment is observed among individuals born in Phnom Penh, as indicated by the larger coefficients presented in Tables B.14, B.15, and B.16 in Chapter 2 Appendix.

Taken together the results of Tables 2.1, B.4, and B.5 suggest the implications of the Pol Pot regime's urban depopulation policy on the educational attainment of the cohort born in urban areas become apparent, particularly for those born in Phnom Penh and 1977. The negative estimated coefficients indicate the distinct educational disadvantage for individuals born in urban districts during the genocide compared to those born in rural areas. Plausibly, the differential effects observed among urban born cohorts between 1975 and 1979 can be attributed to the active involvement of their parents in the conflict, as the zero urban population policy restricted civilians from residing in cities, except for Khmer Rouge leaders, soldiers, and prisoners. In the findings of Ichino and Winter-Ebmer (2004), children born between 1930 and 1939 whose fathers actively participated in World War II experienced significant declines in educational attainment and earnings.

Among the affected cohorts, individuals born in 1977 exhibit the largest and most statistically significant negative effects. This pattern can plausibly be attributed to the severity of conditions during that year. Historical accounts, including those by Kiernan (2002), Vickery (1983), suggest that 1977 marked a turning point in the Khmer Rouge regime. Living conditions deteriorated sharply as internal purges removed earlier local officials and replaced them with more radical and militarized leadership. These new leaders often imposed stricter controls, increased labor demands, and enforced harsher punishments, particularly targeting the "New People." In addition, food shortages worsened as the regime stockpiled supplies in preparation for a possible war with Vietnam. Children born in 1977 were likely exposed to these severe conditions from birth, including during in utero and early infancy, both of which are critical periods for human capital formation. This early-life exposure to extreme deprivation likely contributes to the disproportionately negative outcomes observed for this cohort.

### 2.4.2 Result of estimating wealth index score

Figure 2.6 depicts the evolution of wealth index scores by years of birth between 1965 to 1984. The figure demonstrates that the cohort born during the historical episode exhibits lower levels of wealth compared to other cohorts, with the cohort born in 1977 experiencing the lowest wealth scores, regardless of their district of birth (see the black line). On average, cohorts born in rural areas tend to have lower wealth than those born in urban areas (see brown and blue lines). However, there is a more pronounced decline in wealth during the Pol Pot years for cohorts born in urban areas (see the blue line), particularly among heads of households born in the big three cities (see the green line) and Phnom Penh (see the red line). Consequently, it can be inferred that the cohort born during the Khmer Rouge era generally exhibits lower wealth, especially if they were born in urban areas, and even more so if their birthplace is Phnom Penh.

Figure 2.6: Average wealth index score of heads of households by birth year.

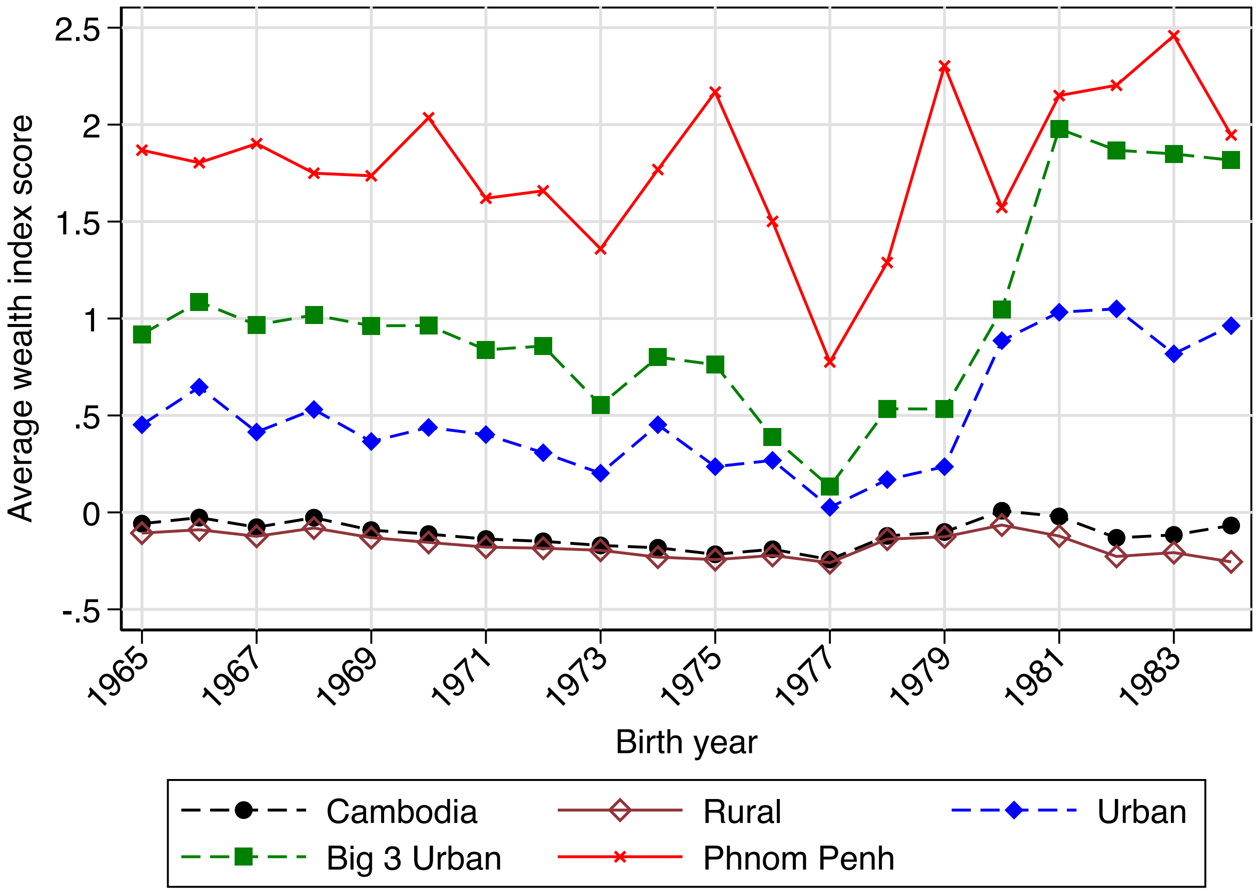



Table 2.2 presents regression results for the wealth index score of the urban cohort born in 1977, having the same structure as Table 2.1. In panel A, using an urban proxy I to interact with the cohort of interest yields statistically significant coefficients at the 5% or 10% level. Conversely, in panel B, which uses urban proxy II to interact with the cohort born in 1977, coefficients are statistically significantly different from zero. Additionally, in panel C, all coefficients are statistically significant at the 1% level, except for column (6), which is significant at the 10% level after controlling for district-specific cubic time trends. Moreover, the coefficients in panels D and E are highly statistically significant. In brief, the urban cohort born in 1977 exhibits lower wealth scores, with even lower scores observed for those born in the big three cities and further low in Phnom Penh. These regression results are consistent with the evidence presented in Figure 2.6. Panels A, B, and C correspond to the blue line, panel D with the green line, and panel E with the red line in Figure 2.6.

Table B.6 focuses on the wealth outcomes of the cohort born during the genocide between 1976-1978 in urban districts, similar to Table 2.2. In panels A and B, where urban proxies I and II are used in the interaction terms, respectively, the coefficients do not attain statistical significance when controlling district-specific cubic time trends (see column (6)). However, the

Table 2.2: Wealth index score and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.02* | -0.02* | -0.03** | -0.02* | -0.02** | -0.03* |
| | (0.011) | (0.012) | (0.012) | (0.010) | (0.011) | (0.013) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel B | | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.00 | -0.00 | -0.01 | -0.00 | -0.01 | -0.01 |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.006) | (0.008) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel C | | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.11*** | -0.12*** | -0.12*** | -0.09*** | -0.11*** | -0.12* |
| | (0.019) | (0.021) | (0.029) | (0.019) | (0.030) | (0.066) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel D | | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.39*** | -0.40*** | -0.40*** | -0.37*** | -0.38*** | -0.35*** |
| | (0.027) | (0.030) | (0.037) | (0.023) | (0.037) | (0.103) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel E | | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -0.88*** | -0.91*** | -0.97*** | -0.81*** | -0.90*** | -0.98*** |
| | (0.111) | (0.115) | (0.129) | (0.104) | (0.141) | (0.221) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 67,603. The urban proxy I$_{77}$ is constructed as $(1 - \frac{PopBorn_{77}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{77}$ is defined as $(1 - \frac{PopBorn_{77}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration. $^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

coefficient in panel C, specifically in column (6), remains significant at the 10% level compared to column (5), which excludes district-specific cubic time trends. Conversely, the results in panels D and E remain consistent regardless of including the district-specific cubic time trends. These outcomes are expected, as the cohort born in urban areas between 1976-1978, despite the levels of statistical significance.

Table B.7 uses the cohort dummy variable for heads of households born between 1975-1979 in the interaction term. The coefficients exhibit no substantial changes in panels A and B, but in column (6), both coefficients become positive for both panels. Meanwhile, panels C, D, and E coefficients are quite similar. Despite being statistically insignificant, the negative signs of the coefficients indicate that the urban cohort born during the Pol Pot era experiences lower wealth than other cohorts or cohort rural born. These findings, however, are consistent with the results in Tables 2.2 and B.6, albeit different cohort dummies are used.

Furthermore, the regression results from females and males separately provide additional evidence supporting the adverse impact of being born in urban areas during the genocide, especially in Phnom Penh, as those in Tables 2.2, B.6, and B.7. The cohort born during the Pol Pot years experiences lower wealth scores if born in urban districts, with even lower scores observed for those born in Phnom Penh, regardless of gender (see Tables B.17, B.18, B.19, B.20, B.21, and B.22 in Chapter 2 Appendix).

To validate the constructed wealth index score, the 2008 census data is used as it contains more asset ownership variables than the 1998 census. The regression results using the 2008 census data are consistent with the findings obtain from the 1998 census, given variations in the asset ownership variables and their statistical significance (see Tables B.23, B.24, and B.25 in Chapter 2 Appendix). Despite the statistical significance, the negative coefficients indicate wealth index score disparities between urban born and rural born cohorts during the Pol Pot regime.

### 2.4.3 Demographic health survey data

Childhood exposure to war and conflict has been found to have a notable impact on physical development, particularly regarding height-for-age, which is commonly used as a proxy for early-life health and nutrition ((Akresh et al., 2012). A large body of research finds that chronic malnutrition and stunting are associated with poorer cognitive development, delayed school enrollment, and fewer years of completed schooling. Nutrition, therefore, plays a crucial role in shaping both educational and wealth outcomes, as health during infancy and early childhood is fundamental for long-term human capital formation (Cunha et al., 2006; Glewwe & Jacoby, 1995; Glewwe & Miguel, 2007). Individuals who were exposed to famine during their early life stages may face disadvantages in adult height, weight, education, and labor market outcomes (Meng & Qian, 2009). Therefore, the 2000 Demographic and Health Survey (DHS) data consolidates the stature measurements of a cohort born during the genocide.[6]

The respondents are women aged 15 to 49 who have been married or lived in a consensual union.

[6]DHS can be retrieved from https://dhsprogram.com

These respondents provide information about their adult height and weight, which is unavailable in census data. Seven observations with heights below 70 centimeters (cm) and above 200 cm are removed as potential outliers. Additionally, four observations with weights below 25 kg and above 150 kg are dropped due to potential outliers. One limitation of the Cambodia DHS dataset is the need for more information about the district of the birthplace, unlike the IPUMS-I dataset. Thus, the regression does not include the birthplace district as the control variable or fixed effect.

Figure B.6 depicts the average height of female individuals born between 1965 and 1984. The cohort born in 1977 appears slightly shorter than others born during and before the Pol Pot era. The downward height trend can be attributed to the ongoing physical development of cohorts born a few years after the Pol Pot period. Similarly, Figure B.7 shows that the cohort born in 1977 has a slightly lower average weight than others. Additionally, Figure B.8 illustrates the evolution of schooling by birth year. Despite the smaller sample size, the trend observed in the DHS data aligns with the pattern observed in the census data (see the black dashed line; for including the 95% confidence interval, see Figure B.9). It indicates that the cohort born in 1977 has, on average, lower levels of education compared to neighboring cohorts. It is important to note that the DHS data lacks information about the district of birthplace but includes the province of residence.

Table B.26 presents the height premium of the cohort born during the Khmer Rouge period. In Panel A, columns (1) - (4) report the estimated heights for the cohort born 1977, with each column varying the control variables and fixed effects. Despite the lack of statistical significance, the corresponding estimates consistently indicate a negative relationship between height and the cohort born in 1977. Columns (5) - (8) present the estimated weight coefficients for cohort 1977, similar to the estimated height coefficients. Although none of the estimated coefficients are statistically significant, the negative coefficients suggest a negative relationship between weight and the cohort born in 1977.

Panel B presents the estimated heights and weights for the cohort born between 1976 and 1978, following a similar structure to Panel A. Despite the lack of statistical significance, both height and weight exhibit a negative relationship with the cohort born during this period. Conversely, in Panel C, columns (1) - (3), a positive relationship is observed between height and the cohort born between 1975 and 1979, with only column (3) showing statistical significance. However, column (4) obtains the negative coefficient when including province-specific cubic time trends. Moreover, the corresponding estimates of weight in columns (5) - (8) indicate a negative relationship, and only column (8) is statistical significance at a 5% level when including the province-specific cubic time trends.

The estimated effects of the cohort born during the genocide do not offer conclusive evidence of the negative impacts of starvation on adult height and weight, similar to the findings of De Walque (2006). However, the absence of conclusive evidence regarding the effect on height and weight in this study highlights the need for additional supporting information about birth districts to validate the hypothesis. Nevertheless, the DHS data does provide valuable insights into the height and weight patterns among the cohort born during the genocide.

## 2.5 Conclusion

This paper examines the long-term effects of being born in urban areas during the Khmer Rouge regime on adult educational attainment and household wealth. Using data from the 1998 Cambodian Census and a generalized difference-in-differences framework, the analysis shows that individuals born in urban districts between 1975 and 1979, particularly in 1977 and in Phnom Penh, experienced significantly lower schooling and wealth levels compared to other cohorts. These results are robust across various alternative measures of urbanization measures and specifications.

While previous studies have primarily focused on the impact of conflict exposure during school-age years, this study contributes new evidence on the enduring disadvantage faced by children born during a period of systemic institutional collapse and forced urban depopulation. The findings indicate that the Khmer Rouge's policy of evacuating urban areas played a central role in shaping long-term human capital outcomes for affected cohorts.

Supplemental analysis using DHS data does not reveal statistically significant differences in adult height or weight for the urban-born cohort. However, this absence of statistical significance does not imply that these individuals were spared malnutrition or early-life health deficits, nor the subsequent consequences for education, wealth, and labor market outcomes. Chronic malnutrition, starvation, and disease were widespread under the regime, and the lasting effects of such conditions on cognitive development and socioeconomic outcomes remain a plausible mechanism behind the adverse impacts observed later in life.

This paper also acknowledges several data limitations, including the absence of direct measures for parental loss, asset destruction, and local recovery capacity. Moreover, the forced relocation of urban residents during the Khmer Rouge regime may have introduced misclassification into the data, as children of urban-origin parents who gave birth in rural areas are recorded as rural-born. This potential measurement error could bias the estimated effects and understate the true disadvantage experienced by the urban-born cohort. These limitations underscore the need for further research to disentangle the specific pathways through which early-life exposure to conflict and displacement affect adult outcomes. For example, recent work Becker et al. (2020) find that descendants of Poles who were forcibly relocated after World War II invested more in education across generations, viewing human capital as a more portable and secure form of investment compared to physical assets. Similarly, Case et al. (2002) show that long-run household income is a powerful determinant of children's health, with lower income leading to faster health deterioration and lower socioeconomic status in adulthood. Such research deepens our understanding not only of how conflict affects human capital development, but also of the long-term social and economic consequences of forced displacement and systemic upheaval.

# Chapter 3

# The Gender Composition in Cambodian Agriculture

## 3.1 Introduction

Rice has long been the dominant crop in Cambodia, with wet rice cultivation shaping the country's agricultural landscape for centuries. Historically, Cambodian rice farming has relied on plow-based methods for land preparation which has contributed to gendered labor divisions. Pryor (1985) characterized rice as a "plow-positive" crop because plowing plays a crucial role in increasing productivity. However, the adoption of plow-based agriculture has also been associated with a persistent gender division of labor. Studies by Boserup (1970), Burton and White (1984), and Alesina et al. (2013) find that societies historically practicing plow-based agriculture tend to have lower female participation in agriculture. This is largely due to the perception that plowing is physically demanding and, therefore, a task suited for men. Over time, such gender norms have shaped agricultural labor structures, contributing to lower female labor force participation even beyond agriculture (Alesina et al., 2013).

While mechanization has transformed Cambodia's agricultural sector in recent decades its impact on gender composition remains unclear. Farmers increasingly adopt tractors, power tillers, water pumps, and threshers (Chhun et al., 2015) and select farming methods and machinery based on land size crop types and land characteristics (Saruth et al., 2014). However, mechanization does not automatically dissolve historical gendered labor divisions. Instead research indicates that the adoption of technology often strengthens male dominance in agriculture by shifting control over mechanized tasks to men while reducing opportunities for female labor (Afridi et al., 2023; Boserup, 1989; Carranza, 2014; Fredriksson & Gupta, 2023; Unnevehr & Stanford, 1985).

Recent reports from the Ministry of Agriculture, Forestry and Fisheries (2022) indicate that crop cultivation remains the most common agricultural activity in Cambodia. However, men outnumber women in this sector, with a similar pattern observed among external laborers, where the number of female workers remains significantly lower than that of males. One key factor contributing to this disparity is the shift in labor roles driven by mechanization, which has allowed men to maintain dominance in the agricultural workforce. Most agricultural machinery is designed by and for men, limiting accessibility for female farmers (Kawarazuka et al., 2018; Rola-Rubzen et al., 2020).

Traditionally, women were responsible for tasks such as harvesting and post-harvesting, but these roles have increasingly shifted to men as agricultural practices evolved (Ministry of Agriculture, Forestry and Fisheries, 2022). Similar patterns have been documented globally. In Korea, the introduction of rice-planting tractors led men to take over transplanting, a task previously dominated by women Boserup (1989). Likewise, in the Philippines and India, mechanization in transplanting, harvesting, and deep tillage reduced female labor demand, contributing to a long-term decline in women's participation in agriculture (see Res (1985) for the Philippines; see Carranza (2014) for India). These findings suggest that historical gender norms continue to shape labor markets, with male-dominated roles persisting even as farming practices modernize.

This paper examines the impacts of rice cultivation on the gender composition of Cambodia's agricultural labor force, arguing that rice, a historically plow-intensive crop, has perpetuated the pattern of gendered labor divisions in agriculture. Despite modernization, labor patterns remain deeply embedded in traditional farming structures. Using farm-level data from the 2019 Cambodia Inter-Censal Agricultural Survey (CIAS) and district-level analysis with IPUMS and GAEZ data, this study analyzes the relationship between rice cultivation and female labor participation in Cambodia.

While past research, such as Alesina et al. (2013), Fan et al. (2024), Hansen et al. (2015), shows that agricultural traditions can shape long-run gender norms across sectors, this study intentionally limits its focus to observable labor force participation in agriculture. By narrowing the scope, I aim to empirically isolate how rice cultivation, rather than broader cultural transmission, affects the gender composition of the agricultural workforce. This sector-specific focus provides a foundation for future research into how historical agricultural practices continue to shape gender norms beyond the agricultural sector in Cambodia.

Empirical results show that farms that allocate more land to rice employ fewer women, even after controlling for farm size, total agricultural employment, and other relevant factors. While overall farm size influences female participation, rice cultivation remains the primary determinant of gender composition in agricultural labor. Comparatively, farms cultivating alternative crops such as cashew, cassava, mango, or maize exhibit higher female labor participation. This pattern suggests that even as agriculture shifts away from draft animals and plows, male labor continues to dominate in rice farming.

At the district level, the relationship between rice productivity and female agricultural labor force remains consistent with farm-level findings, reinforcing the persistent impact of historical labor structures on female employment in agriculture. To address potential reverse causality, this paper employs a two-stage least squares (2SLS) regression using elevation as an instrumental variable (IV). The 2SLS estimates confirm that higher rice productivity is associated with lower female labor participation in agriculture. These results remain robust across multiple specifications, including province-clustered standard errors.

This study highlights that while Cambodia's agricultural sector has modernized, the historical gendered division of labor remains a defining feature of rice cultivation. The findings indicate that agricultural modernization alone does not necessarily lead to a more balanced distribution of female and male agricultural labor, as male-dominated labor patterns continue to define employment in the sector.

The remainder of this paper is organized as follows. Section 3.2 provides a review of the relevant existing literature. Section 3.3 describes the data source. Section 3.4 analyses farm-level outcomes, followed by district-level outcomes in Section 3.5. Section 3.6 concludes.

## 3.2 Literature review

This paper contributes to the broader literature on gender composition in agricultural labor, particularly in the context of rice cultivation in Cambodia. A long-standing body of research shows that plow-based agriculture has historically reinforced gendered labor divisions, with men dominating farming activities due to the physical demands of plowing. These labor patterns have persisted over time, leading to lower female participation in agriculture, particularly in societies where rice cultivation is dominant (Alesina et al., 2013; Boserup, 1970). Hansen et al. (2015) extend this discussion by emphasizing that the longer a society has relied on cereal agriculture, the lower the participation of women in labor force. Their findings suggest that as society moved further from hunting and gathering, agricultural intensification limiting women's long-term economic opportunities.

Building on Boserup's (1970) theory of agricultural intensification, Burton and White (1984) find that plowing is primarily used for soil preparation and planting, tasks that have traditionally been performed by men. However, they also identify a strong relationship between wet rice cultivation and female labor participation in agriculture, particularly in regions with short dry seasons and low cattle presence. Their findings suggest that while plowing reinforces male dominance in agricultural labor, wet rice farming, especially under certain ecological conditions, can create opportunities for higher female labor participation.

Beyond plow-base societies, Fredriksson and Gupta (2023) show that societies with historical reliance on irrigation-intensive agriculture exhibit lower contemporary female labor force participation due to the physical demands of irrigation and male-dominated governance structures. In these societies, men primarily worked in the fields, while women remained closer to the home, engaged in household-related tasks. This historical division of labor has persisted into the present day, further limiting women's participation in agricultural employment. Their findings complement Alesina et al. (2013), who argue that plow-based farming has contributed to the persistence of male dominance in agricultural labor markets.

Technological advancements in agriculture have significantly improved productivity but also transformed labor dynamics. Bustos et al. (2016) provide evidence that labor-saving agricultural technologies, such as genetically engineered soy in Brazil, reduced labor intensity and shifted employment away from agriculture. While their study does not focus on gender, similar labor-saving trends in other agricultural settings have been linked to declines in female labor participation (Carranza, 2014; Paris, 1998). While mechanization may contribute to lower female labor participation, Palacios-Lopez et al. (2017) highlights broader structural factors such as land ownership, household labor composition, and agricultural specialization as key determinants of gendered labor dynamics.

Carranza (2014) provides a key empirical framework for understanding how agricultural conditions influence agricultural laborers. Using soil texture as an instrumental variable, she demonstrates that increased mechanization in Indian agriculture disproportionately reduced female labor force participation. Her findings indicate that a 10% points in the share of loamy soil relative to clayey soils leads to a 5.1% decrease in the share of female laborers and a 7.2% decline in the girl-to-boy ratio, suggesting the demographic consequences of mechanization.

Supporting Carranza (2014) findings, Afridi et al. (2023) show that mechanized tilling has led to a decline in female labor participation in India, primarily by reducing demand for weeding tasks traditionally performed by women. Similarly, Takeshima (2024) examines mechanization trends across seven developing countries and finds that increased adoption of tractors and combine harvesters shifts female labor away from farm-based activities toward non-farm employment. Kawarazuka et al. (2018) and Rola-Rubzen et al. (2020) emphasize that one reason for this trend is that most agricultural tools and equipment are designed without taking women's physiques into account, limiting women's accessibility and increasing their displacement from traditional farming roles. While these studies highlight how mechanization influences labor allocation, they do not explicitly focus on gender composition in rice farming.

In the context of Asian rice farming, Paris (1998) documents that mechanization, particularly the introduction of direct seeding, herbicides, mechanical threshers, and rice mills, has disproportionately displaced female laborers. Women, who were once heavily involved in transplanting, weeding, and post-harvest processing, have seen their roles diminished as these tasks have either been taken over by machines or reallocated to men who operate mechanized equipment. While mechanization has reduced physical labor burdens for some, it has also reinforced gendered labor disparities by shifting control over agricultural technology to men.

While these studies establish a clear relationship between mechanization and gendered labor shifts, they do not specifically examine whether rice farming itself remains a key determinant of female labor participation in agriculture. This study addresses this gap by investigating whether rice cultivation in Cambodia, a historically plow-intensive crop, has continued to strengthen male labor dominance despite modernization.

The paper also relates to literature on the persistence of rice farming effects, particularly from a gender roles perspective. Historically, rice farming has played a significant role in shaping social norms and cultural practices, particularly in regions where it has been dominant for centuries. Talhelm and English (2020) indicate that, historically, rice farming regions in China have strong social norms, community ties, and collective work ethics compared to wheat farming locales. This is because wet rice cultivation required shared irrigation networks and coordinated labor efforts. However, Talhelm and Oishi (2018) argue that rice farming does not always result in collectivism, even though it required cooperation, because different societies may have developed different solutions to rice farming challenges.

Even in the modern era, the cultural influence of historical rice farming persists. Fan et al. (2024) find that former rice farming regions in China continue to exhibit cultural norms that influence people's preferences, ranging from strong family ties to the prevalence of family-controlled firms. This persistence of agricultural effects on social and gender norms parallels the long-term influence of historical plow-based agriculture. Alesina et al. (2011) explore how historical plough practices are shaping fertility preferences and find a negative relationship between plough use and fertility rates today. The reason is that the economic role of women and children in plow agriculture, diminished the need for larger families. Similarly, Alesina et al. (2018) find that cultural norms from societies that used ploughs in the past still exist today and are associated with a higher male-to-female ratio in modern times. These findings suggest that just as historical plow agriculture contributed to lasting gendered labor structures, traditional rice farming also played a role in shaping persistent patterns of female labor force participation.

## 3.3 Data

### 3.3.1 Data description

This study utilizes data from the first Cambodia Inter-Censal Agriculture Survey (CIAS) National Institute of Statistics of the Kingdom of Cambodia (2021), which is nationally representative for Cambodia and is used as the core dataset for analysis. Additionally, several other datasets are used for validity checks, including the Integrated Public Use Microdata Series - International (IPUMS-I) , the Global Agro-Ecological Zones (GAEZ) project, topographic data from the EarthEnv project, global climate and weather data from WorldClim, and the Caloric Suitability Index (CSI).

### 3.3.2 Farm-level data

The 2019 farm-level microdata is the first Inter-Censal Agriculture Survey (CIAS), nationally representative for Cambodia. The survey was conducted from the beginning of July 2018 to the end of June 2019. The data covers 25 provinces, excluding 8 urban districts.[1] The sample is collected only from households with agricultural holdings and activities involving growing crops, raising livestock and poultry, and fishing activities. There are about 16,000 agricultural households in the sample after excluding the 8 urban districts. However, for the purpose of this study, only households involved in growing crops are selected, so the sample reduces to $15,983$.

The sample contains agricultural households who either have only homelots or parcels, or have both homelots and parcels. About 61% of all agriculture products are mainly for home consumption,

[1]Six districts are in Phnom Penh: Chamkar Mon, Doun Penh, Prampir Meakkakra, Tuol Kouk, Ruessei Kaev, and Chbar Ampov. One district is Preah Sihanouk city in Preah Sihanouk province and one core urban district is Siem Reap city in Siem Reap province.

the rest are for sale. There are 42 crops grown in Cambodia, and the most popular crops are aromatic paddy, non-aromatic paddy, mango, and banana. Table C.1 in ?? presents summary statistics for the top ten crops grown by agricultural holdings that hired external and occasional workers in the studied sample. On average, 82% of households grow rice, followed by 7% for cassava, and 3% for cashew. The remaining seven crops, each grown by less than 2% of households, include maize, soybeans, sugarcane, and rubber.

The data also contains details about agricultural household members working on the farm and paid or unpaid external agricultural workers, including both males and females. It also includes information on whether these workers are involved in crop cultivation, raising livestock, or aquaculture. However, I select only those engaged in crop production, which reduces the total number of observations to $5,132$. Additionally, the external workers can be categorized as full-time, part-time, or occasional workers. Most occasional workers are involved in crop production, with 53% being female. Meanwhile, the data does not contain geospatial information on the location of each household's agricultural holdings, nor does it include wage and salary information.

Figure 3.1 depicts the farms that cultivate rice tend to employ fewer female workers. In contrast, Figure 3.2 illustrates that farms growing alternative crops, such as cashew, cassava, mango, or maize, have a higher proportion of female workers.

Figure 3.1: Average fraction of female workers and rice dummy.

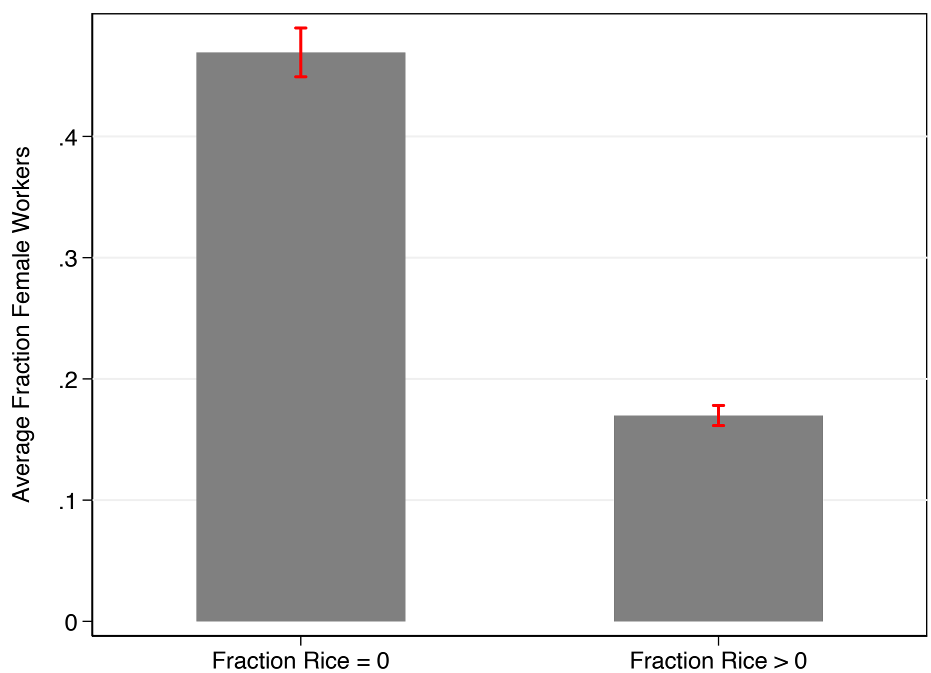


Figure 3.2: Average fraction of female workers and dummy variables of other top four crops.

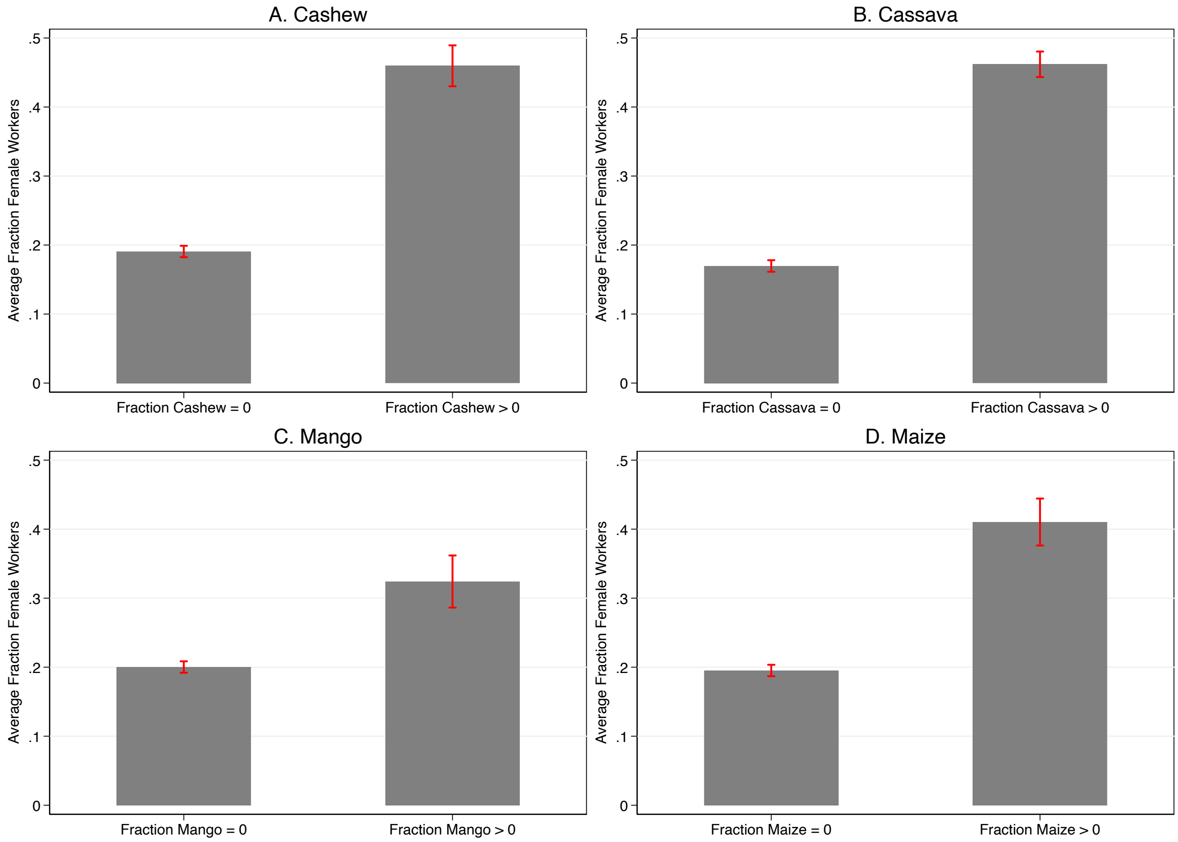



Table 3.1 presents summary statistics for the estimated fraction of female workers in agriculture. There are few points to note. First, I selected only farms growing crops and hiring external or occasional workers for this study, so the total observations is reduced to 5,132. On average, 82% of all farms grow rice. Second, in the sample, I do not distinguish between homelots and parcels as long as their purpose is growing crops. Lastly, the average family schooling is calculated by adding together the years of schooling completed by each household member. The number of years of schooling is determined based on the midpoint of the highest level of education completed by each individual. For instance, 3 years are assigned for completing primary education, 8 for secondary education, 11 for high school, 18 for a bachelor's/master's/PhD degree, and 14 for other technical diplomas. Thus, average family schooling, on average, is 4.65 years.

Panel A in Table C.2 provides summary statistics for the main variables based on farms grow some rice, while Panel B provides those that do not. The data shows that, on average, 95% of all agricultural holdings grow rice, but they employ only about 17% of female workers. Additionally, among the 93% of rice-farming holdings, the average total farm size is approximately 2.6 hectares per holding. Households that grow rice tend to have more female members than those cultivating other crops.

Table 3.1: Summary statistics of farm-level main variables.

| | Mean | Std. Error | Min. | Max. | Median | Obs. |
|---|---|---|---|---|---|---|
| Fraction female workers | 0.21 | 0.30 | 0.00 | 1.00 | 0.00 | 5,132 |
| Fraction rice | 0.82 | 0.35 | 0.00 | 1.00 | 1.00 | 5,132 |
| Total farm size | 3.33 | 8.45 | 0.00 | 330.00 | 1.50 | 5,132 |
| Log total farm size | 0.38 | 1.21 | -7.60 | 5.80 | 0.41 | 5,132 |
| Total number of workers | 7.67 | 11.86 | 1.00 | 194.00 | 4.00 | 5,132 |
| Log total number of workers | 1.53 | 0.91 | 0.00 | 5.27 | 1.39 | 5,132 |
| Average family schooling | 4.65 | 2.69 | 0.00 | 18.00 | 4.25 | 5,132 |
| Fraction female in households | 0.30 | 0.20 | 0.00 | 1.00 | 0.25 | 5,132 |

*Notes:* The Cambodia Inter-Censal Agriculture Survey (CIAS) 2019 collected data on 42 different crops. Farms were categorized into two types: those used for rice paddy cultivation and those for non-rice paddy crops. The fraction rice is calculated as the total land area used for rice paddy divided by the total farm size, while the fraction non-rice paddy is computed as one minus the fraction rice. The dataset includes 5,132 farm households that employed occasional and external workers across various provinces. The fraction of female workers refers to the proportion of occasional and external workers who are female. Total farm size is measured in hectares.

### 3.3.3 District-level data

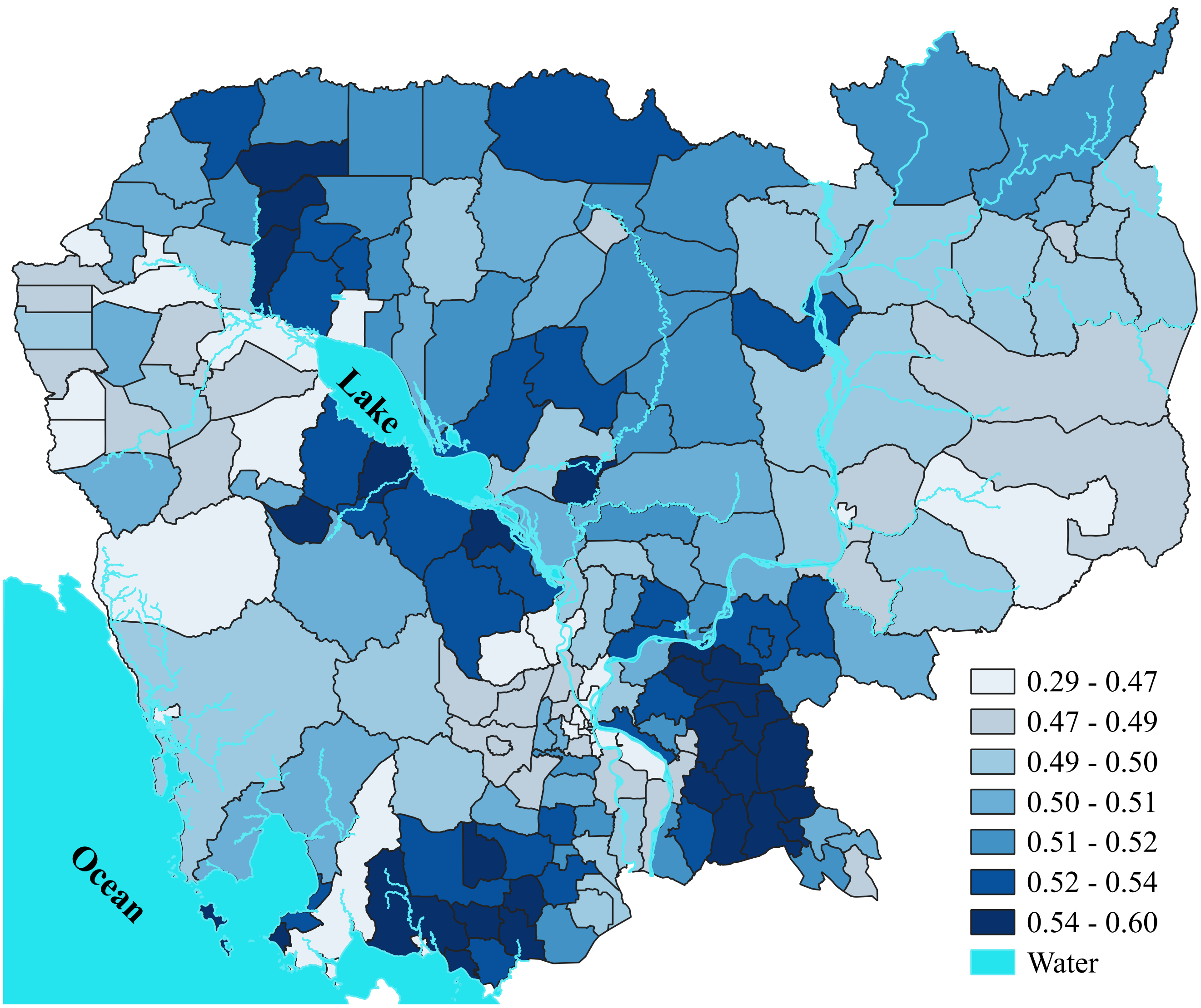
#### 3.3.3.1 Census data

This paper uses district-level data from the Integrated Public Use Microdata Series - International (IPUMS-I) for the 2019 census, which covers 10 percent of the total population and is nationally representative. The dataset includes geospatial information at both the province and district levels for household samples. It provides detailed individual demographic data, including occupation status, years of schooling, and highest education level attained. However, one limitation of the dataset is the lack of information on household agricultural holdings and the use of external and/or occasional agricultural workers. To address this, I restrict the sample to individuals who report their sector, activities, or products as related to the cultivation of non-perennial crops. The final analytic sample includes $395,104$ individuals aged 15 to 69, drawn from a total IPUMS sample of $1,522,877$ individuals.

While a large share of Cambodia's population resides in rural areas, not all rural residents are actively engaged in crop agriculture, nor are all included in occupationally coded census responses. This filtered sample thus represents a conservative but occupationally meaningful estimate of the crop-related agricultural workforce used in the district-level analysis.

Figure 3.3 displays the fraction of females among agricultural labor by district. There are four ranges based on the proportion of female participation rate in agricultural activities. Districts with darker red shading indicate a higher fraction of agricultural workforce, while lighter green color represents districts with the lowest proportion of female laborers. The data show that female agricultural labor participation is highest around Tonle Sap Lake (Lake), in the southern regions, and in the northwest. This distribution suggests that women play a signification role in agricultural production, particularly in low elevation areas, which also tend to be in provinces with high population density.

Figure 3.3: Fraction females among agricultural labors by district.



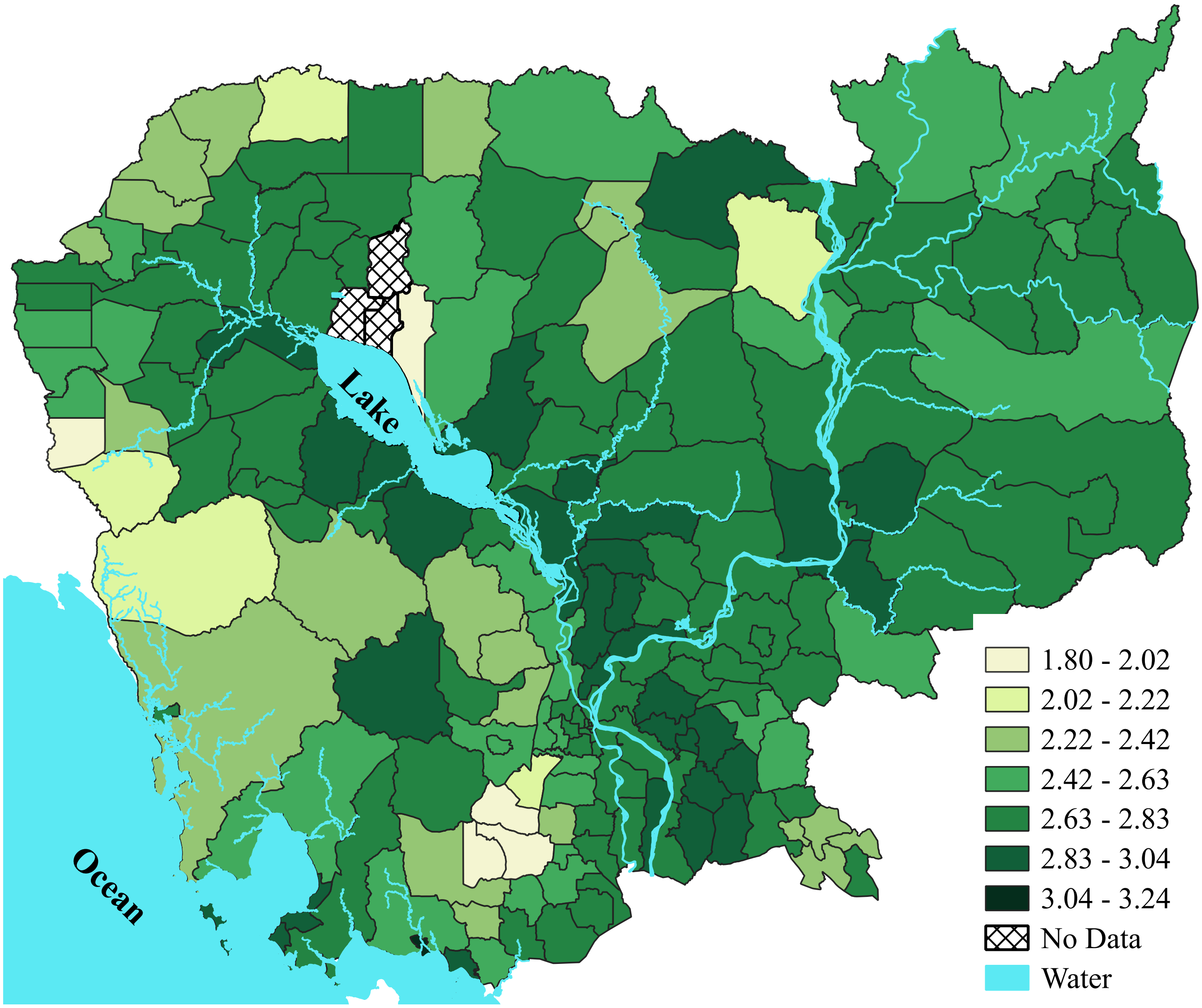

#### 3.3.3.2 Actual and potential yield data

I use data from the Global Agro-Ecological Zones (GAEZ) project, developed by Food and Agriculture Organization (FAO) and Applied Systems Analysis (IIASA), which provides detailed information on climate, soil, and terrain for agriculture.[2] GAEZ combines high-resolution geographic data (5 arc-minute grids, roughly $9 \times 9$ $km$) with crop-specific agronomic models to estimate potential yields. Each grid cell includes key factors such as soil quality (e.g., depth, fertility), climate (e.g.,

[2]GAZE data portal link: https://gaez.fao.org

temperature, precipitation), and terrain (e.g., elevation, slope).

Figure 3.4: Average actual rice yield by district in 2010.




Yield calculations consider different water supply scenarios (irrigated, rainfed) and various farming inputs, from traditional subsistence farming to fully mechanized practices. Low-level inputs involve subsistence farming with no fertilizers or chemicals, intermediate inputs involve moderate mechanization and chemical use, and high-level inputs involve fully mechanized farming with optimal resource use. This paper focuses on potential yields under low-input management and rainfed conditions, with the baseline reference years from 1961 to 1990.

GAEZ also calculates actual yields and production. For 2000 and 2010, yields were mapped using harvested area, production, and yield data for 26 crops, based on 3-year averages 1999-2001 and 2009-2011, respectively. Rainfed and irrigated land were included, with production values based on 2000 international prices. Since Food and Agriculture Organization Statistics (FAOSTAT) provides only general national data, GAEZ applies a "downscaling" method, using geospatial data and land characteristics, soil, climate, and terrain, to distribute national data across grid cells. This allows for localized and fine-scale analysis of agricultural production. This paper uses rainfed actual rice yields in 2010 for the main results.

Figure 3.4 illustrates the mean rice productivity across districts in Cambodia. Darker green shades represent higher rice productivity, whereas lighter shades indicate lower productivity levels. However, three districts in Siem Reap province, namely Krong Siem Reap, Prasat Bakong, and Banteay Srei, do not have data on rice yield. The map also highlights that rice cultivation is mostly concentrated around water bodies, as wet rice farming depends on sufficient water availability.

#### 3.3.3.3 Geo-characteristic and other data

The elevation data used in this paper is sourced from EarthEnv, a project supported by NCEAS, NASA, NSF, and Yale University.[3] EarthEnv project is a collaboration between biodiversity scientists and remote sensing experts to create standardized 1 $km$ resolution data layers for monitoring and modeling biodiversity and ecosystems. By utilizing global elevation models the 250 $m$ GMTED2010 and 90 $m$ SRTM 4.1dev, EarthEnv provides topographic data on variables including elevation, slope, aspect, eastness, northness, roughness, and more. These variables are available at various spatial aggregations, ranging $1, 5, 10, 50$, and 100 $km$. This paper uses the mean elevation at 1 $km$ aggregation from the SRTM 4.1 dev model.

The Figure 3.5 illustrates the elevation levels across districts in Cambodia. The map shows that much of the country is low-lying with vast areas at 0-50 meters, particular around the Tonle Sap Lake (Lake) and the Mekong River Basin, where agriculture is concentrated. Moderate elevations between 50-500 meters appear mainly in the northeast and southwest. In contrast, higher elevations exceeding 500 meters are around in mountainous regions in the southwest, north and east. These areas are characterized by rugged terrain and dense forests. Interestingly, the lower elevation areas also correspond to the regions with the highest rice productivity (see Figure 3.4).

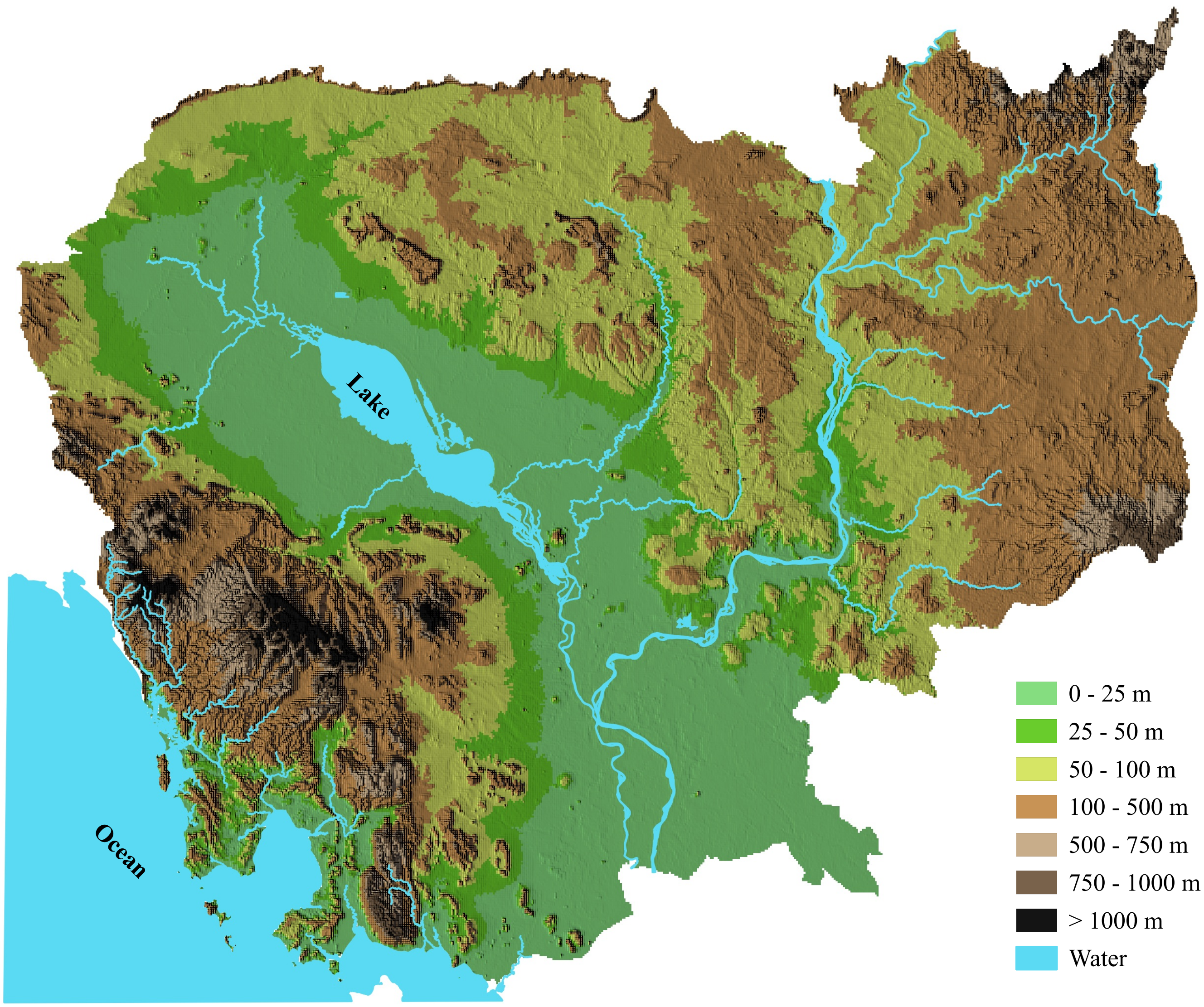
For the temperature and precipitation, the paper uses the 30-second (1 $km^2$) resolution data from the global climate and weather database, WorldClim.[4] WorldClim is a global database providing high-resolution climate data for land areas from 1970 to 2000. The dataset includes variables such as temperature, precipitation, solar radiation, vapor pressure, and wind speed, all interpolated from 9,000 to 60,000 weather stations. In regions with low station density, satellite data is used to enhance temperature predictions. The data is available at multiple spatial resolutions: 30 seconds (approximately 1 $km^2$), 2.5 minutes (about 21 $km^2$), 5 minutes (about 85 $km^2$), and 10 minutes (about 340 $km^2$).

To mitigate the potential influence of agriculture on social and economic factors in the regression, this paper employs the Caloric Suitability Index (CSI) developed by Galor and Özak (2015).[5] The

[3]Elevation data is available through http://www.earthenv.org
[4]WorldClim data is available through https://www.worldclim.org/data/index.html
[5]CSI data can be retrieved from https://ozak.github.io/Caloric-Suitability-Index/

CSI measures potential agricultural output, expressed in calories, based on crops available both before and after 1500 CE. The data is provided at both the grid cell level ($5' \times 5'$) and the country level. For each grid cell, the CSI offers four estimates: the maximum and average potential caloric yields, based on crops available in the pre-1500 and post-1500 periods. This allows for a comparative analysis of agricultural productivity across different historical eras.

Figure 3.5: Elevation.




To integrate all the data, this paper uses year-specific second-level geography 2019 data from IPUMS-I. These year-specific variables offer more detailed geographic information than spatially harmonized variables, as they do not adjust for administrative boundary changes over time. For confidentiality purposes, geographic units with populations under $20,000$ are aggregated.[6]

[6]GIS boundary files can be retrieved from https://international.ipums.org/international/gis.shtml

## 3.4 Farm-level analysis

In this section, the paper presents the baseline farm-level results, examining the effects of rice cultivation on female participation in agricultural employment. The farm-level analysis employs a simple regression model with province fixed effects, and standard errors are clustered at the province level. Section 3.4.1 outline empirical approach, Section 3.4.2 reports the main estimates for female workers in agricultural employment, and Section 3.4.3 examines the robustness of the results by using maize cultivation as an alternative control and considering other crop types as the baseline.

### 3.4.1 Empirical framework

Farm-level data provides specific details such as farm size, crop choices, and labor decisions, which are likely to have a direct impact on female labor participation. To explore the relationship between rice cultivation and female agricultural labor, I estimate the effect of the fraction of land used for rice cultivation on the fraction of females in the agricultural workforce, using province fixed effects. The regression is estimated via an ordinary least square (OLS) using the following equation

$$Y_{ip} = \beta_1 R_{ip} + \beta_2 X_{ip} + \lambda_p + \varepsilon_{ip}, \tag{3.1}$$

where $Y_{ip}$ is the fraction female among agricultural labor in farm $i$ in province $p$. $R_{ip}$ is the fraction of land used for rice cultivation of farm $i$ in province $p$, and represents a farm household's decision to hire workers based on their sex.[7] It is used to examine employment practices and potential biases in hiring decisions within an agricultural setting. $\beta_1$ is the primary interest in this study, and captures the relationship between female workers and the rice cultivation through the share of land used for rice cultivation.

$X_{ip}$ is a vector of control total farm size, total number of agricultural workers, average family years of schooling, and fraction of female household member. These controls help provide a clearer picture of the relationship between rice cultivation and female labor. For example, the total farm size accounts for the effects of the overall size of farm land in the particular province because a larger farm may have different labor dynamics, more machinery, or growing more multi crops compared to a smaller one.

The control of total agricultural worker can ensure that the relationship between rice cultivation and female employment is not simply because a lot of rice farms also happen to have more workers. Furthermore, the average schooling of the farm household's family, rather than solely the household head's schooling, helps better explain whether the observed relationship is due to the type of work women do in more educated farm family or due to rice cultivation itself. This approach avoids the limitations of relying on headship as a stratifying variable, as mentioned by Quisumbing (1996), who suggests that focusing only on the household head's education ignores the roles and decisions

[7] The fraction of land used for rice cultivation is the total land used for rice paddy divided by the total farm size.

of other family members, particularly women, who may have their own economic activities and farm their own plots within male-headed households. At the same time, control for farm household female member would ensures that the relationship rice cultivation and female workers is not because some households have more women, but the rice cultivation itself affects female worker. Plausibility, a province has more women it might have more female labor participation.

$\lambda_P$ is province fixed effects. The estimation also adjusts standard errors clustering within province.

### 3.4.2 Farm-level results

Table 3.2 shows the benchmark result for examining the relationship between fraction of female among agricultural workers and the fraction of land cultivated rice against non-rice.

Table 3.2: Fraction females and rice: Non-rice as base.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Fraction rice | -0.35*** | -0.35*** | -0.24*** | -0.24*** | -0.24*** | -0.16*** | -0.16*** |
| | (0.010) | (0.012) | (0.012) | (0.012) | (0.012) | (0.014) | (0.040) |
| Log total farm size | | 0.01 | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** |
| | | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) | (0.008) |
| Log total number of workers | | | 0.16*** | 0.16*** | 0.16*** | 0.16*** | 0.16*** |
| | | | (0.004) | (0.004) | (0.004) | (0.004) | (0.017) |
| Average family schooling | | | | -0.00*** | -0.00*** | 0.00 | 0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | -0.00 | 0.03* | 0.03 |
| | | | | | (0.018) | (0.017) | (0.022) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.17 | 0.17 | 0.37 | 0.37 | 0.37 | 0.46 | 0.46 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary least square (OLS) regressions. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are categorized into two types: those used for rice paddy, and those used for non-rice paddy. Fraction rice is defined as the total land used for rice paddy divided by the total farm size. Fraction non-rice paddy equals one minus fraction rice.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Column 1 presents a negative association between female workers and rice cultivation. It indicates that farms that allocate a greater share of their land to rice tend to hire fewer female workers compared to farms that allocate more land to other crops. A 10 percentage points in the share of land used for rice cultivation corresponds to a 3.5 percentage points decrease in employment of female workers.

In column 2, the logarithm of the total farm size is included as a control to account for the overall scale of the farm, as farm size may influence the decisions of hiring female workers and grow rice. However, since the coefficient of farm size is small, close to zero, and not statistically significant, hiring more or less female workers does not driven by the farm size. The coefficient of the main variable of interest remains unchanged, indicating that rice cultivation itself primarily influences the gender composition of agricultural labor.

Column 3, introduces the log of the total number of workers as an additional control to capture the effect of the size of the agricultural employment on female workers. Controlling for both farm size and workers size allows for isolating the effect of rice cultivation on female labor. The positive and highly significant coefficient suggests that a larger employment is associated with a higher fraction of female workers. Moreover, the coefficient for farm size changes from negative to positive and becomes statistically significant when controlling for the total number of workers. Meanwhile, the coefficient for rice cultivation remains robust, even though its magnitude decreases from $-0.35$ to $-0.24$. This implies that rice farming is less favorable for female workers compared to other crops, depending on the scale of farms and size of employment. Clearly, lager farms may hire fewer female workers because of potentially increased mechanization or reliance on male workers for certain tasks.

In column 4, I look at the implication of employment opportunity choice by controlling the average years of schooling of farmers' families. The fact that families have higher education levels, traditional gender roles may shift with more employment options for women outside agriculture with fewer women participation in labor-intensive agricultural activities. Hence, it may influence both the decision to participate in agriculture and the type of farming practices adopted by the families. Family with higher education may prefer less labor-intensive crops or adopt advanced farming techniques that require fewer manual laborers, particularly female workers. While the negative estimate is statistically significant, the coefficient is close to zero, suggesting that household education has a minor negative effect on female workers. However, since other coefficients remain unchanged in both magnitude and significance, it implies that household education attainments may influence the absolute level of the fraction of female workers. Nonetheless, it does not alter the relationship between female workers and rice cultivation.

In column 5, for better understanding if the fraction of female workers is influenced by the household's internal gender dynamics besides agricultural factors, fraction of female members in the household is added as additional control. This is because household structure and gender composition may affect both labor supply and allocation within families. More female members in household might have a greater supply of potential female labor, and thus hiring fewer external female workers. However, since the coefficient is negative and not statistical significant, it suggests that the observed effect on female workers arises from agricultural factors rather than the household gender composition and education.

Moreover, in column 6, province fixed effects are included to better understand the regional dynamics. After controlling for province fixed effect, the absolute value of the coefficient for fraction of land used for rice cultivation reduces from $-0.24$ to $-0.16$ but remains highly significant. It indicates that rice farming might rely more on family labor, especially women, rather than external paid workers in the region with more share of rice farms. This is because the coefficient of fraction of female members in household changed from negative to positive and significant at 10% level. The relationship between farm size, total workers, and female workers are consistence across different provinces as the coefficients unchanged and remain significant.

Lastly, in column 7, standard errors are clustered at the province level to account for within-province correlations, and the results remain robust. This confirms that the relationships observed in Table 3.2 are not sensitive to provincial clustering and maintain their statistical significance across specifications.

Additionally, to build on the results in Table 3.2, Tables C.4 and C.5 in **??** provide weighted regressions using the total number of workers and total farm size, respectively. These analyses further explore the relationship between rice cultivation and female workers using different implications. The findings from Tables C.4 and C.5 support the conclusions from Table 3.2, confirming that rice cultivation is associated with hiring fewer female workers. Consequently, the evidence presented in Table 3.2 strongly suggests that rice cultivation has a direct effect on the gender composition of agricultural labor. Thus, it tends to hire fewer female workers relative to farm size and the overall agricultural workforce. This suggests that farms with higher female employment might focus on non-rice crops or other agricultural activities. The nature of labor intensive for rice cultivation may favor male workers due to physical demands, which could explain the lower employment of female workers.

Furthermore, other mechanisms may also explain the direct effect of rice cultivation on female agricultural labor that this paper could not directly observe from the data. In Western Kenya, women's participation in agriculture is often constrained by structural barriers such as limited access to land, credit, and farming inputs, which significantly reduce their productivity and ability to engage in commercial farming (Diiro et al., 2018). More broadly, in Africa, structural transformation has led to a reallocation of female labor away from agriculture and toward the service sector, driven by economic development, urbanization, and the decline of home production burdens (Dinkelman & Ngai, 2022). Similarly, in Cambodia, between 2011 and 2016, women rapidly transitioned from agriculture into wage employment in manufacturing and services, particularly in the garment sector, as the economy moved away from traditional agricultural employment (Gavalyugova & Cunningham, 2020). In Lao PDR, where the modernization of lowland rice farming has pushed women towards non-farm activities, while men retain control over rice farming. Women in Lao PDR adopt new agricultural technologies at similar rates to men but more likely to abandon them if the benefits are too small. Women also report greater difficulties in accessing farming markets, particularly multiple buyers when selling rice, which can limit their ability to engage in commercial rice farming (Moglia et al., 2020). These structural constraints may further limit female labor opportunities in rice farming, exacerbating gender disparities in the agricultural sector.

### 3.4.3 Farm-level robustness checks

To ensure that the benchmark results are not sensitive to how rice farming is measured, the share of land used for rice cultivation is replaced with a rice dummy variable. This dummy is defined as 1 if a farm grows any amount of rice, and 0 otherwise. This approach ensures that the relationship between female workers and rice cultivation is consistent under a different specification. Table C.6 in **??** presents the regression results using the rice dummy, and they are similar to the benchmark results in Table 3.2. This suggests that the presence of rice farming itself affects female employment independently of the quantity of rice grown. Even small amounts of rice cultivation can lead to fewer female workers, it reinforces the finding that rice farming genuinely reduces opportunities for women to participate in paid agricultural labor.

The relationship between rice cultivation and female workers might be alter by its share relative to all other crops combined. To isolate the impact of each crops on female workers, I estimate the effects of the top 10 crops most commonly grown by agricultural holdings, one at a time, to identify which crop most potentially affects female employment.[8] This method help clarify whether the effect of rice cultivation on female workers is independent of other crops. Table C.7 reports the results for each crop, showing that only the coefficients for rice and oil and fiber are negative. These negative coefficients indicate a negative association between these two crops and female employment. However, since the coefficient for oil and fiber is not statistically significant and is very close to zero, its effect on female workers is less reliable than that of rice. Consequently, rice cultivation demonstrates a strong adverse relationship with female employment. Therefore, Table C.7 results support the robustness of the findings in Table 3.2.

Following the regression results in Table 3.2, Table C.8 in **??** is conducted a similar analysis with a twist of comparing the share of land used for rice cultivation against the share of land used for maize cultivation. This approach aims to determine whether the lower hiring of female workers is specific to rice farming or if the effect change when compared to maize. By doing so, the regression provides a clearer understanding of how these two crops differ in their impact on female labor participation in agriculture.

The results in Table C.8 indicate that the effect of rice cultivation on fraction female paid workers remain negative and statistically significant consistent with the findings in Table 3.2. However, the coefficients sightly decrease (e.g., from $-0.35$ to $-0.28$, or from $-0.16$ to $-0.13$, in column 1 and 7, respectively,) when comparing Table 4 (table above) and Table 5 (this table). This reduction suggests that while rice cultivation continues to have the adverse effect on female workers, its impact is slightly weaker when directly compared to maize cultivation rather than non-rice cultivation. Additionally, the drop in the number of observations from 5132 in Table 4 to $4,662$ in Table C.8 reflects a smaller sample size as the analysis focuses on farms that cultivate both rice and maize cultivation. Overall, the similar results of Table 3.2 and Table C.8 demonstrate that rice cultivation consistently has the negative impact on female workers, and the effect is directly from rice farming,

[8] The top ten growing crops are rice paddy, cassava, cashew, maize, mango, fruit-bearing vegetables, rubber, vegetable tree/leaf/flower, ool and fiber, and sugarcane.

regardless of whether the comparison is made against non-rice crops in general or maize specifically.

Furthermore, Table C.9 in **??** presents regression results comparing rice cultivation to fruit-bearing vegetables. Both crops are seasonal staples; however, rice requires intensive manual labor for planting, transplanting, and harvesting, whereas fruit-bearing vegetables need flexible labor for pruning, pest control, and multiple harvests throughout the growing season. Comparing these two crops helps to understand how different crop types affect female employment, given their varying labor demands and intensities. The results in Table C.9 are consistent with those in Table 3.2, suggesting that rice cultivation tends to employ fewer female workers than fruit-bearing vegetables.

Table C.10 in **??** replicates the regression from Table 3.2, but with an added control for the fraction of land used for maize, keeping all other variables the same. By including both rice and maize in the regression, the effects of each crop on female workers is considered independently. This approach helps clarify whether rice or maize has the greater impact on hiring decisions, given that rice cultivation tends to require more physical strength and seasonal cycle, while maize typically does not. The coefficients of rice and maize cultivation are $-0.36$ and $-018$, respectively, and highly statistical significant (see column 1). The negative coefficients indicate that as more land is allocated to either rice or maize cultivation, there is a significant decrease in female workers. However, rice cultivation has the stronger negative effect on female employment than maize cultivation does, given its larger coefficient magnitude. The effect of rice is more pronounced possibly because it demands more physical labor traditionally associated with male workers, whereas maize's cultivation may involve different gender roles and have less gender-biased labor demand.

Moreover, columns 6 and 7, in Table C.10, show the coefficients for maize became positive and no statistical significant holding others constant when control province fixed effect, regardless of robust or clustering standard error are used. The changed in coefficient after including province fixed effect suggests that the previously observed negative relationship between female workers and maize cultivation is likely driven by the unobserved provincial characteristics, such as crop suitability and climate, non-agricultural job opportunities, income, education and skills, migration, or season rather than maize cultivation itself. The opposite effects of rice and maize cultivation on female employment imply that rice cultivation has the dominant negatively effect on hiring female workers.

The findings suggest that rice cultivation consistently has a strong negative effect on female employment in agriculture. Therefore, the evidence presented in this section strongly supports the results in Table 3.2, indicating that rice farming directly influences the gender composition in agriculture and tends to employ fewer women.

Before presenting the district-level analysis, it is important to note a key difference in reported female labor shares across the two datasets. In the farm-level CIAS data, women comprise approximately 21% of hired agricultural workers (see Table C.2), whereas in the district-level IPUMS sample, the share of women in agricultural occupations is closer to 50% (see Table C.3).

This discrepancy reflects differences in measurement and scope. The CIAS dataset captures external and occasional labor hired by crop-producing households, whereas IPUMS records self-reported primary occupations, including unpaid family workers and subsistence farmers. Accordingly, CIAS represents demand-side hiring decisions, while IPUMS captures the broader labor supply in agriculture. Rather than being inconsistent, these two figures offer complementary insights into the gendered structure of agricultural labor in Cambodia.

## 3.5 District-level analysis

To complement and extend the findings from the farm-level analysis, this paper examines whether the relationship between rice cultivation and female labor participation persists at the district level. This broader analysis adopts a supply-side perspective, drawing on individuals' self-reported occupations rather than farm-level labor demand. The key independent variable is the log of average district-level rice yield, and the dependent variable is the share of agricultural workers who are female. The analysis begins with an ordinary least squares (OLS) regression to assess the baseline association, and then employs a two-stage least squares (2SLS) regression to address potential endogeneity.

Section 3.5.1 outlines empirical framework for both OLS and 2SLS approaches. Section 3.5.2 reports the OLS and 2SLS estimates, respectively, for female among agricultural occupation.

### 3.5.1 Empirical framework

#### 3.5.1.1 Ordinary Least Square approach

This section estimates the effect of rice productivity on the gender composition of the agricultural labor force using an OLS regression at the district level. The estimating equation is as follows

$$Y_{dp} = \beta_1 R_{dp} + \beta_2 X_{dp} + \delta_p + \varepsilon_{dp}, \tag{3.2}$$

where $Y_{dp}$ denotes the fraction females among agricultural labor force in district $d$ within province $p$. $R_{dp}$ is the log of average rice yield in 2010. The coefficient $\beta_1$ captures the association between rice yield and female labor participation in agriculture. The vector $X_{dp}$ includes several control variables to account for demographic and geographic heterogeneity across districts: the fraction of women in the total workforce, the share of the population engaged in agriculture, the log of population density, log precipitation, log temperature, log caloric suitability index, log distance to Phnom Penh, and log distance to the nearest river.

Controlling for these district-level characteristics helps isolate the association between rice yield and the gender composition of the agricultural labor force from potential confounding factors. For example, more populous districts may offer greater non-agricultural employment opportunities, reducing reliance on agricultural work, particularly for women. Controlling for population density helps account for the influence of urbanization on labor allocation. Similarly, controlling for the distance to the nearest river captures geographic variation in irrigation potential and crop selection, both of which can influence labor demand. The specification also includes province fixed effects $\delta_p$ to control for unobserved provincial-level heterogeneity, and standard errors are clustered at the province level to account for intra-provincial correlation.

Table C.3 provides summary statistics for the district-level sample. On average, women engaged about 50% of agricultural labor force, and about 52% of the total workforce. The average share of the population employed in agriculture is relatively low at 29%. The log of rice yield has a mean of 0.97 with a standard deviation of 0.09.

#### 3.5.1.2 Two-Stage Least Square approach

To further refine the analysis and address potential endogeneity between rice productivity and female labor participation, the 2SLS regression is used. In the first stage, the potentially endogenous variable, log rice yield $R_{dp}$ in district $d$ in province $p$, is instrumented using elevation $E_{dp}$. The first-stage equation is specified as

$$R_{dp} \quad = \quad \alpha_1 E_{dp} + \alpha_2 X_{dp} + \lambda_p + \eta_{dp}, \tag{3.3}$$

where $X_{dp}$ is a vector of control variables consistent with those in Equation (3.2), and $\lambda_p$ denotes province fixed effect. Standard errors are clustered at the province level to account for potential correlation of outcomes within provinces.

In the second stage, the predicted values of rice yield from the first stage $\widehat{R}_{dp}$ are used to estimate their effect on the share of women among agricultural labor, $Y_{dp}$, using the following specification

$$Y_{dp} \quad = \quad \beta_1 \widehat{R}_{dp} + \beta_2 X_{dp} + \lambda_p + \varepsilon_{dp}, \tag{3.4}$$

The coefficient $\beta_1$ captures the causal effect of rice productivity on female participation in the agricultural labor force. As in the first stage, $X_{dp}$ includes all relevant controls, and $\lambda_p$ accounts for unobserved provincial-level heterogeneity.

Elevation serves as a plausible instrumental variable in this context because it is strongly correlated with rice productivity. Lowland areas in Cambodia, particularly those surrounding the Tonle Sap lake, are more suitable for rice cultivation due to better water availability, flatter terrain, and more fertile soils. Figures 3.5 and 3.4 show that rice farming is concentrated in these low-elevation regions, supporting the instrument's relevance through a strong first-stage relationship.

As a fixed geographic feature, elevation is unlikely to be influenced by social, economic, or policy-driven forces that may affect district-level labor market outcomes. It does not change over time or respond to household or market decisions. Thus, any influence of elevation on female participation in agriculture would plausibly operate through its effect on crop suitability, rather than through direct social or economic channels.

3.3 provides further support for this interpretation. Districts with higher female participation in agricultural labor tend to be more densely populated, but this pattern does not appear to systematically align with elevation. Moreover, the analysis controls for a range of potentially confounding district-level characteristics, such as population density, the share of agricultural employment, precipitation, and distance to rivers and Phnom Penh, helps to address observable sources of bias and reduces the risk of omitted variable confounding.

Although the exclusion restriction cannot be formally tested in a just-identified IV setting, the inclusion of comprehensive controls strengthens the case for treating elevation as an exogenous source of variation. Moreover, Angrist and Kolesár (2024) demonstrate that conventional 2SLS inference remains reliable under typical empirical conditions, particularly when the first-stage relationship is strong and the estimated coefficient has the expected sign.

This identification strategy provides a more credible estimate of how rice productivity relates to the gender composition of agricultural labor across Cambodian districts.

### 3.5.2 District-level results

#### 3.5.2.1 District-level OLS results

Table 3.3 presents the main results from the OLS regression. Column 1 reports the relationship between the fraction of women in agricultural occupations and rice productivity, measured as the log of average district-level rice yield. The specification controls for the female share of the total workforce, the share of the population employed in agriculture, and the log of population density.

The coefficient on rice productivity is negative ($-0.04$) and statistically significant, suggesting that districts with higher rice yields tend to have a lower share of women in agricultural labor. In contrast, both the overall female workforce share and the agricultural employment share are positively associated with female participation in agriculture. This pattern implies that even in areas where women are more active in the labor force or where agriculture is more prevalent, higher rice productivity is still linked to reduced female presence in farming.

This pattern may reflect how higher rice yields are associated with more intensive farming practices that shift the nature of labor demand. For example, high-yield districts could be more mechanized or rely more heavily on physically demanding or time-intensive tasks, factors that may reduce the relative share of women in agricultural work. While not directly measured, these mechanisms offer a plausible explanation for the observed association between rice productivity and the gender composition of agricultural labor.

Columns 2-4 introduce additional geographic controls, including the log distance from each district to Phnom Penh and to the nearest river. The inclusion of these variables does not substantially change the results. The negative relationship between rice yield and female agricultural participation remains consistent and statistically significant, supporting the robustness of the initial finding.

Columns 5-8 replicate the specifications in Columns 1-4 but apply clustering of standard errors at the province level. While the estimated coefficients remain negative and similar in magnitude, the significance levels decline, especially in Columns 5 and 6. This suggests that clustering standard errors at the province level increases the standard errors. The loss of statistical significance points to the presence of unobserved variation within provinces that robust standard errors in earlier columns may not fully capture.

Table 3.3: Fraction females among agricultural occupation and log actual rice yield in 2010.

| | Dependent variable is the fraction female among agricultural occupation | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log actual rice yield in 2010 | -0.04* | -0.04** | -0.05** | -0.04** | -0.04 | -0.04 | -0.05* | -0.04* |
| | (0.021) | (0.020) | (0.020) | (0.021) | (0.029) | (0.025) | (0.023) | (0.025) |
| Fraction female in workforce | 0.77*** | 0.75*** | 0.76*** | 0.75*** | 0.77** | 0.75** | 0.76** | 0.75** |
| | (0.231) | (0.222) | (0.226) | (0.229) | (0.325) | (0.339) | (0.345) | (0.347) |
| Fraction population in agriculture | 0.11*** | 0.11*** | 0.10*** | 0.11*** | 0.11*** | 0.11*** | 0.10*** | 0.11*** |
| | (0.030) | (0.032) | (0.032) | (0.034) | (0.029) | (0.030) | (0.030) | (0.030) |
| Log population density | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.006) | (0.007) | (0.007) | (0.007) | (0.004) | (0.004) | (0.004) | (0.004) |
| Log precipitation | | -0.00 | -0.01 | -0.00 | | -0.00 | -0.01 | -0.00 |
| | | (0.035) | (0.035) | (0.036) | | (0.050) | (0.048) | (0.049) |
| Log temperature | | 0.03 | 0.05 | 0.10 | | 0.03 | 0.05 | 0.10 |
| | | (0.256) | (0.260) | (0.264) | | (0.326) | (0.337) | (0.356) |
| Log caloric suitability index | | -0.02 | -0.02 | -0.01 | | -0.02 | -0.02 | -0.01 |
| | | (0.106) | (0.108) | (0.110) | | (0.120) | (0.122) | (0.125) |
| Log distance of each district to Phnom Penh | | | 0.01 | 0.01 | | | 0.01 | 0.01 |
| | | | (0.010) | (0.011) | | | (0.013) | (0.013) |
| Log distance of each district to river | | | | 0.01 | | | | 0.01 |
| | | | | (0.005) | | | | (0.004) |
| Province fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Cluster | Cluster | Cluster | Cluster |
| $R^2$ | 0.59 | 0.59 | 0.59 | 0.60 | 0.59 | 0.59 | 0.59 | 0.60 |
| Number of observations | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 |

*Notes:* Ordinary Least Squares (OLS) Regressions. Robust standard errors are in parentheses. In columns 5 through 8, standard errors are clustered by province. The unit of observation is the district.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Overall, the OLS results indicate a consistent negative association between rice productivity and the female share of agricultural labor. However, the sensitivity of the estimates to clustering raises concerns about omitted variable bias and unobserved province-level factors that may jointly effects both rice yields and female labor supply. Although the analysis controls for several relevant factors, the possibility of reverse causality, where female labor availability influences crop choices or productivity, remains a potential concern

To address these concerns, the next section turns to a two-stage least squares (2SLS) approach that uses elevation as an instrument for rice productivity, allowing for a more reliable estimation of the causal relationship.

#### 3.5.2.2 District-level 2SLS results

Table C.12 presents the first-stage regression results, examining the relationship between the log of elevation (used as an instrumental variable) and the log of actual rice yield (the endogenous variable). Across all specifications, the coefficient on log elevation is consistently negative and statistically significant, ranging from $-0.06$ to $-0.08$. These results are robust to the inclusion of different control variables and province fixed effects, supporting the instrument's relevance.

The first-stage F-statistics range from 8.02 to 9.47 in columns 1 through 4 (with robust standard errors), and from 7.89 to 14.84 in columns 5 through 8 (with province-clustered standard errors). These values raise some concern about weak instruments, particularly in the earlier specifications. As noted by Staiger and Stock (1997), a first-stage F-statistic below 10 may indicate a weak instrument, potentially biasing second-stage estimates. Stock and Yogo (2005) formalized this rule of thumb by providing critical values. They show that an F-statistic below 10.27 implies more than a 10 percent relative bias in 2SLS estimates.

Nevertheless, recent work by Angrist and Kolesár (2024) cautions against the mechanical application of the F-statistic threshold of 10. They argue that strict pretesting based on this rule can distort inference and potentially lead to the unwarranted rejection of otherwise valid instruments. In this analysis, the Cragg–Donald Wald F-statistics exceed 38 across all specifications, suggesting that the instrument is sufficiently strong. Furthermore, the Stock–Wright LM S-statistics do not reject the null of instrument validity in any specification, lending additional support to the use of elevation as a credible instrument.

Taken together, these additional diagnostics suggest that elevation is sufficiently strong for use as an instrument, even if some first-stage F-statistics fall slightly below conventional thresholds. Furthermore, Angrist and Kolesár (2024) propose a "sign screening" approach, which treats the instrument as valid if the first-stage coefficient has the expected sign, even when F-statistics are modest. In this case, the consistently negative and significant coefficients further support the use of elevation as an exogenous predictor of rice productivity.

Table 3.4 shows the second-stage 2SLS regression results, estimating the causal effect of log rice yield on the female share of agricultural labor. Several diagnostic tests are reported in the table to assess the instrument's strength and validity. The Kleibergen-Paap Wald F-statistics range from 29.84 to 39.69, indicating that the instrument is strong and unlikely to suffer from weak identification. The Stock-Wright LM test also supports the validity of the instrument by confirming the strong correlation between elevation and rice yield. Furthermore, the endogeneity test p-values range from 0.02 to 0.15, suggesting that OLS estimates may be biased, which justifies the use of the IV approach.

Columns 1 to 4 report the main coefficient estimates for log rice yield, which range from $-0.12$ to $-0.19$ and are statistically significant at the 5% to 1% levels. These effects are larger in absolute value than the OLS estimates. Thus, it suggests that OLS may understate the true causal effect due to endogeneity.

Columns 5 to 8 replicate the same specifications using standard errors clustered at the province level. In columns 6 to 8, the coefficients remain negative and statistically significant, supporting the robust inverse relationship between rice productivity and female agricultural labor. The coefficient in column 5, however, remains statistical insignificant similar to its OLS in the same specification in Table 3.3. This variation suggests that statistical significance may depend on how standard errors are adjusted.

Overall, the 2SLS results indicate that higher rice productivity reduces the share of female workers in agriculture. The estimated effects are consistently negative and statistically significant in most specifications and are larger in magnitude than the OLS estimates. These findings suggest that OLS may be downward biased due to endogeneity. Moreover, the 2SLS results consistent with finding from farm-level analyses and provide the same conclusion that rice cultivation, the crop historically associated with plow use, reduces female agricultural participation.

## 3.6 Conclusion

This study provides empirical evidence on how rice cultivation influences the gender composition of Cambodia's agricultural labor force. Using farm-level and district-level data, the findings reveal that farms allocating more land to rice cultivation tend to employ fewer female workers, contributing to persistent gender divisions in agricultural labor. A two-stage least squares (2SLS) regression, using elevation as an instrumental variable, confirms a causal relationship between higher rice productivity and lower female agricultural employment.

These results are close to Carranza (2014), who demonstrates that agricultural conditions significantly affect female agricultural laborers. Similar to her findings, this study shows that farms cultivating rice, a historically male-dominated crop continue to employ fewer female workers. The persistence of rice farming deepens gender roles and perpetuates the existing gender composition of the Cambodian agricultural sector. This persistence of rice farming is similar to Boserup (1970) and Alesina et al. (2013), who argue that historical plow-based agricultural societies are associated with lower female participation in agriculture. Similarly, Fredriksson and Gupta (2023) find that irrigation-intensive agriculture, much like plow-based farming, has historically contributed to reduced female labor force participation.

Table 3.4: Fraction females among agricultural occupation and log actual rice yield in 2010.

| | Dependent variable is the fraction females among agricultural occupation | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log actual rice yield in 2010 (instr.) | -0.12** | -0.17*** | -0.19*** | -0.19*** | -0.12 | -0.17* | -0.19** | -0.19** |
| | (0.060) | (0.061) | (0.066) | (0.066) | (0.073) | (0.085) | (0.080) | (0.081) |
| Fraction female in workforce | 0.78*** | 0.72*** | 0.74*** | 0.73*** | 0.78** | 0.72** | 0.74** | 0.73** |
| | (0.217) | (0.215) | (0.221) | (0.224) | (0.316) | (0.328) | (0.336) | (0.339) |
| Fraction population in agriculture | 0.11*** | 0.11*** | 0.11*** | 0.11*** | 0.11*** | 0.11*** | 0.11*** | 0.11*** |
| | (0.029) | (0.031) | (0.031) | (0.032) | (0.026) | (0.029) | (0.029) | (0.029) |
| Log population density | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.006) | (0.007) | (0.007) | (0.007) | (0.004) | (0.004) | (0.004) | (0.004) |
| Log precipitation | | 0.01 | 0.01 | 0.01 | | 0.01 | 0.01 | 0.01 |
| | | (0.027) | (0.028) | (0.029) | | (0.036) | (0.037) | (0.038) |
| Log temperature | | 0.23 | 0.29 | 0.32 | | 0.23 | 0.29 | 0.32 |
| | | (0.250) | (0.262) | (0.265) | | (0.321) | (0.340) | (0.351) |
| Log caloric suitability index | | -0.07 | -0.07 | -0.06 | | -0.07 | -0.07 | -0.06 |
| | | (0.102) | (0.105) | (0.106) | | (0.117) | (0.120) | (0.122) |
| Log distance from district to Phnom Penh | | | 0.02* | 0.02 | | | 0.02* | 0.02 |
| | | | (0.010) | (0.010) | | | (0.011) | (0.011) |
| Log distance from district to river | | | | 0.00 | | | | 0.00 |
| | | | | (0.005) | | | | (0.005) |
| Province fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Cluster | Cluster | Cluster | Cluster |
| Kleibergen-Paap Wald F statistic | 33.37 | 37.27 | 30.08 | 30.00 | 29.84 | 39.69 | 30.43 | 31.94 |
| Stock-Wright LM S: P value | 0.02 | 0.00 | 0.00 | 0.00 | 0.09 | 0.02 | 0.00 | 0.00 |
| Endogenous test: P value | 0.12 | 0.02 | 0.01 | 0.01 | 0.15 | 0.07 | 0.03 | 0.02 |
| $R^2$ | 0.19 | 0.14 | 0.11 | 0.11 | 0.19 | 0.14 | 0.11 | 0.11 |
| Number of observations | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 |

*Notes:* Two-Stage Least Squares (2SLS) Regressions. The endogenous variable is the logarithm of actual rice yield in 2010, instrumented using the logarithm of elevation. Robust standard errors are in parentheses. In columns 5 through 8, standard errors are clustered by province. The unit of observation is the district.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Despite the ongoing modernization of agriculture, some tasks in rice production have become less labor-intensive; however, the deep-rooted association between "plow-positive" crops and male-dominated labor roles remains evident. The persistence of these historical labor structures suggests that modernization alone is not necessarily associated with greater female participation in rice cultivation. In contrast, farms cultivating non-plow-intensive crops tend to employ a higher proportion of female workers, suggesting that agricultural labor structures continue to reflect the historical dominance of men in plow-based farming systems.

These findings contribute to a broader understanding of gendered labor dynamics in agriculture and highlight the persistence of historical agricultural practices in shaping contemporary employment patterns. However, this study has several limitations. First, the data does not capture specific occupational roles within farming, limiting an in-depth analysis of how gender divisions manifest in specific agricultural tasks. Second, the absence of wage data prevents an examination of potential gender wage gaps. Lastly, as this study focuses on Cambodia, its findings may not be directly generalization to other agricultural economies with different historical and structural contexts.

Future research should explore the mechanisms driving gender disparities in rice farming, including land ownership structures and labor market conditions. Additionally, investigating the impact of changing agricultural technologies on gender roles across different crop types could offer further insights into the evolving structure of Cambodia's agricultural labor market.

# Bibliography

Acemoglu, D., Hassan, T. A., & Robinson, J. A. (2011). Social structure and development: A legacy of the Holocaust in Russia. *Quarterly Journal of Economics*, *126*(2), 895–946.

Afridi, F., Bishnu, M., & Mahajan, K. (2023). Gender and mechanization: Evidence from Indian agriculture. *American Journal of Agricultural Economics*, *105*(1), 52–75.

Akresh, R., Bhalotra, S., Leone, M., & Osili, U. O. (2012). War and stature: Growing up during the Nigerian civil war. *American Economic Review*, *102*(3), 273–277.

Akresh, R., & de Walque, D. (2008). Armed conflict and schooling: Evidence from the 1994 Rwandan genocide. *World Bank Policy Research Working Paper*, (4606).

Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford economic papers*, *58*(3), 450–474.

Alesina, A., Giuliano, P., & Nunn, N. (2011). Fertility and the Plough. *American Economic Review*, *101*(3), 499–503.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, *128*(2), 469–530.

Alesina, A., Giuliano, P., & Nunn, N. (2018). Traditional agricultural practices and the sex ratio today. *PLOS ONE*, *13*(1), 1–14. https://doi.org/10.1371/journal.pone.0190510

Angrist, J., & Kolesár, M. (2024). One instrument to rule them all: The bias and coverage of just-ID IV. *Journal of Econometrics*, *240*(2), 105398. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.12.012

Ashraf, Q. H., Weil, D. N., & Wilde, J. (2013). The effect of fertility reduction on economic growth. *Population and Development Review*, *39*(1), 97–130.

Attanasio, O., Meghir, C., & Nix, E. (2020). Human capital development and parental investment in India. *The Review of Economic Studies*, *87*(6), 2511–2541.

Banister, J., & Johnson, P. (1993). *After the nightmare: The population of Cambodia*. Center for International Research.

Barham, T. (2012). Enhancing cognitive functioning: Medium-term effects of a health and family planning program in Matlab. *American Economic Journal: Applied Economics*, *4*(1), 245–273.

Becker, S. O., Grosfeld, I., Grosjean, P., Voigtländer, N., & Zhuravskaya, E. (2020). Forced migration and human capital: Evidence from post-WWII population transfers. *American Economic Review*, *110*(5), 1430–1463.

Bhalotra, S., Karlsson, M., Nilsson, T., & Schwarz, N. (2022). Infant health, cognitive performance, and earnings: Evidence from inception of the welfare state in Sweden. *Review of Economics and Statistics*, *104*(6), 1138–1156.

Blattman, C., & Miguel, E. (2010). Civil war. *Journal of Economic Literature*, *48*(1), 3–57.

Bloom, D. E., & Canning, D. (2008). Global demographic change: Dimensions and economic significance. *Population and Development Review*, *34*, 17–51.

Bloom, D. E., Canning, D., Rosenberg, L., & Ghani, E. (2011). Demographic change and economic growth. *edited by EJAZ GHANI*, 31.

Boserup, E. (1970). Male and female farming systems. In *Woman's role in economic development* (pp. 15–31). 175 Fifth Avenue, New York, New York 10010, USA, St. Martin's Press, Inc.

Boserup, E. (1989). Population, the status of women, and rural development. *Population and Development Review*, *15*, 45–60.

Bühler, M., & Madestam, A. (2024). State repression, exit, and voice: Living in the shadow of Cambodia's killing fields. CEPR Discussion Paper No. 18815, CEPR Press. https://cepr.org/publications/dp18815

Burton, M. L., & White, D. R. (1984). Sexual division of labor in agriculture. *American Anthropologist*, *86*(3), 568–583.

Bustos, P., Caprettini, B., & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, *106*(6), 1320–1365.

Carranza, E. (2014). Soil endowments, female labor force participation, and the demographic deficit of women in India. *American Economic Journal: Applied Economics*, *6*(4), 197–225.

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American economic review*, *92*(5), 1308–1334.

Chandler, D. (2018). *A history of Cambodia*. Routledge.

Chen, Y. J., Li, L., & Xiao, Y. (2022). Early-life exposure to tap water and the development of cognitive skills. *Journal of Human Resources*, *57*(6), 2113–2149.

Chhun, C., Bora, B., & Sothy, E. (2015). *Effect of labour movement on agricultural mechanisation in Cambodia*. Cambodia Development Resource Institute. https://cdri.org.kh/public/storage/pdf/wp107e_1617792510.pdf

Clayton, T. (1998). Building the new Cambodia: Educational destruction and construction under the Khmer Rouge, 1975-1979. *History of Education Quarterly*, *38*(1), 1–16.

Collier, P. (2008). *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press, USA.

Collier, P., & Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford Economic Papers*, *56*(4), 563–595.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American economic review*, *97*(2), 31–47.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, *1*, 697–812.

Currie, J., & Hyson, R. (1999). Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. *American economic review*, *89*(2), 245–250.

Davila, R. L., McCarthy, A. S., Gondwe, D., Kirdruang, P., & Sharma, U. (2014). Water, walls, and bicycles: Wealth index composition using census microdata. *Journal of Demographic Economics*, 1–42.

Davila, R. L., McCarthy, A. S., Gondwe, D., Kirdruang, P., & Sharma, U. (2022). Water, walls, and bicycles: Wealth index composition using census microdata. *Journal of Demographic Economics*, *88*(1), 79–120.

Davis, D. R., & Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, *92*(5), 1269–1289.

De Walque, D. (2006). The socio-demographic legacy of the Khmer Rouge period in Cambodia. *Population Studies*, *60*(2), 223–231.

De Walque, D., & Verwimp, P. (2010). The demographic and socio-economic distribution of excess mortality during the 1994 genocide in Rwanda. *Journal of African Economies*, *19*(2), 141–162.

Deaton, A., & Paxson, C. (2000). Growth and saving among individuals and households. *Review of Economics and Statistics*, *82*(2), 212–225.

Denton, F. T., & Spencer, B. G. (1973). A simulation analysis of the effects of population change on a neoclassical economy. *Journal of Political Economy*, *81*(2, Part 1), 356–375.

Diiro, G. M., Seymour, G., Kassie, M., Muricho, G., & Muriithi, B. W. (2018). Women's empowerment in agriculture and agricultural productivity: Evidence from rural maize farmer households in western Kenya. *PLOS ONE*, *13*(5), e0197995.

Dinkelman, T., & Ngai, L. R. (2022). Time use and gender in Africa in times of structural transformation. *Journal of Economic Perspectives*, *36*(1), 57–80.

Easterly, W., Gatti, R., & Kurlat, S. (2006). Development, democracy, and mass killings. *Journal of Economic Growth*, *11*(2), 129–156.

Ebihara, M., & Kiernan, B. (1993). A Cambodian village under the Khmer Rouge, 1975–1979. *Genocide and democracy in Cambodia: the Khmer Rouge, the United Nations and the international community, edited with an introduction by Ben Kiernan*, (41), 51–63.

Fan, C., Jiang, M., & Zhang, B. (2024). Beyond cultural norms: How does historical rice farming affect modern firms' family control? *Economica*, *91*(363), 770–808. https://doi.org/https://doi.org/10.1111/ecca.12519

FAO and IIASA. (2024). Global Agro-Ecological Zones version 4 (GAEZ v4) [Accessed July 30, 2024]. http://www.fao.org/gaez/

Feyrer, J. (2007). Demographics and productivity. *Review of Economics and Statistics*, *89*(1), 100–109.

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fredriksson, P. G., & Gupta, S. K. (2023). Irrigation and gender roles. *Journal of Development Economics*, *163*, 103076.

Galor, O., & Özak, Ö. (2015). Land productivity and economic development: Caloric suitability vs. agricultural suitability. Brown University working paper. http://papers.ssrn.com/abstract=2625180

Galor, O., & Özak, Ö. (2016). The agricultural origins of time preference. *American Economic Review*, *106*(10), 3064–3103.

Gavalyugova, D., & Cunningham, W. (2020). Gender analysis of the cambodian labor market [License: CC BY 3.0 IGO]. World Bank, Washington, DC. http://hdl.handle.net/10986/34201

Glewwe, P., & Jacoby, H. G. (1995). An economic analysis of delayed primary school enrollment in a low income country: The role of early childhood nutrition. *The review of Economics and Statistics*, 156–169.

Glewwe, P., & Miguel, E. A. (2007). The impact of child health and nutrition on education in less developed countries. *Handbook of development economics*, *4*, 3561–3606.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The lancet*, *369*(9555), 60–70.

Hansen, C. W., Jensen, P. S., & Skovsgaard, C. V. (2015). Modern gender roles and agricultural history: The Neolithic inheritance. *Journal of Economic Growth*, *20*, 365–404.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, *24*(3), 411–482.

Heston, A., Summers, R., & Aten, B. (2012). Penn world table version 7.1. centre for international comparisons of production. Income and prices. University of Pennsylvania, Philadelphia.

Heuveline, P. (1998). 'Between one and three Million': Towards the demographic reconstruction of a decade of Cambodian history (1970-79). *Population Studies*, *52*(1), 49–65.

Heuveline, P. (2003). Mortality and fertility interactions: New insights from recent population dynamics in Cambodia. Population Research Center, NORC. University of Chicago.

Heuveline, P. (2015). The boundaries of genocide: Quantifying the uncertainty of the death toll during the Pol Pot regime in Cambodia (1975–79). *Population Studies*, *69*(2), 201–218.

Holck, S. E., & Cates Jr, W. (1982). Fertility and population dynamics in two Kampuchean refugee camps. *Studies in Family Planning*, 118–124.

Huang, W., Lei, X., Ridder, G., Strauss, J., & Zhao, Y. (2013). Health, height, height shrinkage, and SES at older ages: evidence from China. *American Economic Journal: Applied Economics*, *5*(2), 86–121.

Humphreys, J. (2015). Education premiums in Cambodia: Dummy variables revisited and recent data. *Econ Journal Watch*, *12*(3), 339–345.

Ichino, A., & Winter-Ebmer, R. (2004). The long-run educational cost of World War II. *Journal of Labor Economics*, *22*(1), 57–87.

Islam, A., Ouch, C., Smyth, R., & Wang, L. C. (2016). The long-term effects of civil conflicts on education, earnings, and fertility: Evidence from Cambodia. *Journal of Comparative Economics*, *44*(3), 800–820.

Islam, A., Ouch, C., Smyth, R., & Wang, L. C. (2017). The intergenerational effect of Cambodia's genocide on children's education and health. *Population and Development Review*, *43*(2), 331–353.

Karra, M., Canning, D., & Wilde, J. (2017). The effect of fertility decline on economic growth in Africa: A macrosimulation model. *Population and Development Review*, *43*, 237–263.

Kawarazuka, N., Prain, G., Forsythe, L., Mayanja, S., Mudege, N. N., Babini, C., & Polar, V. (2018). Gender in agricultural mechanization: Key guiding questions. *International Potato Center.*

Kesternich, I., Siflinger, B., Smith, J. P., & Winter, J. K. (2014). The effects of World War II on economic and health outcomes across Europe. *Review of Economics and Statistics*, *96*(1), 103–118.

Khieu, S. (2004). *Cambodia's Recent History and the Reasons Behind the Decisions I Made* [Chapter 7, pp. 53–64]. Phnom Penh, Ponleu Khmer Print. & Pub. House.

Kiernan, B. (1996). *The Pol Pot regime: Race, power, and genocide in Cambodia under the Khmer Rouge, 1975-79*. Yale University Press.

Kiernan, B. (2002). *The pol pot regime: Race, power, and genocide in Cambodia under the Khmer Rouge, 1975-79*. Yale University Press.

Kiernan, B. (2004). *How Pol Pot came to power: Colonialism, nationalism, and communism in Cambodia, 1930-1975*. Yale University Press.

Lindstrom, D. P., & Berhanu, B. (1999). The impact of war, famine, and economic decline on marital fertility in Ethiopia. *Demography*, *36*(2), 247–261.

Maccini, S., & Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American economic review*, *99*(3), 1006–1026.

Markevich, A., Naumenko, N., & Qian, N. (2021). *The causes of Ukrainian famine mortality, 1932-33* (Working Paper No. 29089). National Bureau of Economic Research. https://doi.org/10.3386/w29089

Meng, X., & Gregory, R. G. (2002). The impact of interrupted education on subsequent educational attainment: A cost of the Chinese Cultural Revolution. *Economic Development and Cultural Change*, *50*(4), 935–959.

Meng, X., & Qian, N. (2009). The long term consequences of famine on survivors: Evidence from a unique natural experiment using China's great famine [BER Work Pap 14917].

Meng-Try, E. (1981). Kampuchea: A country adrift. *Population and Development Review*, 209–228.

Merrouche, O. (2011). The long term educational cost of war: Evidence from landmine contamination in Cambodia. *Journal of Development Studies*, *47*(3), 399–416.

Miguel, E., & Roland, G. (2011). The long-run impact of bombing Vietnam. *Journal of development Economics*, *96*(1), 1–15.

Ministry of Agriculture, Forestry and Fisheries. (2022). *Gender Mainstreaming Policy and Strategic Framework in Agriculture 2022–2026*. Phnom Penh, Cambodia, Ministry of Agriculture, Forestry and Fisheries.

Moglia, M., Alexander, K. S., Larson, S., Greenhalgh, G., Thammavong, P., Thephavanh, M., Case, P., et al. (2020). Gendered roles in Agrarian Transition: A Study of Lowland Rice Farming in Lao PDR. *Sustainability*, *12*(13), 1–20.

National Institute of Statistics of the Kingdom of Cambodia. (2021). Cambodia Inter-Censual Agricultural Survey 2019 (CIAS 2019), public use dataset (March 2021). https://nada.nis.gov.kh/index.php/catalog/?page=1&ps=15

Neupert, R. F., & Prum, V. (2005). Cambodia: Reconstructing the demographic stab of the past and forecasting the demographic scar of the future. *European Journal of Population/Revue européenne de Démographie*, *21*(2-3), 217–246.

Nobles, J., Frankenberg, E., & Thomas, D. (2015). The effects of mortality on fertility: Population dynamics after a natural disaster. *Demography*, *52*(1), 15–38.

Özak, Ö. (2015). Caloric Suitability Index - Data (v1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14714917

Palacios-Lopez, A., Christiaensen, L., & Kilic, T. (2017). How much of the labor in African agriculture is provided by women? *Food Policy*, *67*, 52–63. https://doi.org/https://doi.org/10.1016/j.foodpol.2016.09.017

Paris, T. R. (1998). Technology and policy needs of poor women in asian rice farming. *Gender, Technology and Development*, *2*(2), 187–218. https://doi.org/10.1080/09718524.1998.11909893

Pryor, F. L. (1985). The invention of the plow. *Comparative Studies in Society and History*, *27*(4), 727–743.

Quisumbing, A. R. (1996). Male-female differences in agricultural productivity: Methodological issues and empirical evidence. *World development*, *24*(10), 1579–1595.

Res, L. (1985). Changing labor allocation patterns of women in rice farm households: A rainfed rice village, Iloilo province, Philippines. In I. R. R. Institute (Ed.), *Women in rice farming: Proceedings of a conference on women in rice farming systems, the international rice research institute, p.o. box 933, Manila, Philippines, 26-30 september 1983* (pp. 91–115). Aldershot, Hants, England, Gower.

Riano, J. F., & Valencia Caicedo, F. (2024). Collateral damage: The legacy of the secret war in Laos. *Economic Journal*, *134*(661), 2101–2140.

Robinson, N. et al. (2014a). EarthEnv-DEM90: High-Resolution Global Digital Elevation Data [Accessed from http://www.earthenv.org/DEM.html].

Robinson, N., Regetz, J., & Guralnick, R. P. (2014b). Earthenv-dem90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused aster and srtm data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 57–67. https://doi.org/10.1016/j.isprsjprs.2013.09.002

Rola-Rubzen, M. F., Paris, T., Hawkins, J., & Sapkota, B. (2020). Improving gender participation in agricultural technology adoption in Asia: From rhetoric to practical action. *Applied Economic Perspectives and Policy*, *42*(1), 113–125.

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Heimann, Q., Lee, J., & Merrill, N. (2025). Integrated Public Use Microdata Series, International: Version 7.6 [dataset]. *Minneapolis, MN*, IPUMS. https://doi.org/10.18128/D020.V7.6

Saruth, C., Lytour, L., & Sinh, C. (2014). Status and prospect of agricultural mechanization in Cambodia. *Economic and Social Commission for Asia and the Pacific, the United Nation, Bangkok, Thailand*.

Schmitt-Grohé, S., & Uribe, M. (2005). Optimal fiscal and monetary policy in a medium-scale macroeconomic model: Expanded version. National Bureau of Economic Research Cambridge, Mass., USA.

Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1975). *The methods and materials of demography* (Vol. 2). US Department of Commerce, Bureau of the Census.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586. https://doi.org/10.2307/2171753

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). Cambridge University Press. http://papers.ssrn.com/abstract=1734933

Takeshima, H. (2024). Agricultural mechanisation and gendered labour activities across sectors: Micro-evidence from multi-country farm household data. *Journal of Agricultural Economics*, *75*(1), 425–456. https://doi.org/https://doi.org/10.1111/1477-9552.12564

Talhelm, T., & English, A. S. (2020). Historically rice-farming societies have tighter social norms in China and worldwide. *Proceedings of the National Academy of Sciences*, *117*(33), 19816–19824. https://doi.org/10.1073/pnas.1909909117

Talhelm, T., & Oishi, S. (2018). How rice farming shaped culture in Southern China. In A. Uskul (Ed.), *Socioeconomic environment and human psychology* (pp. 53–76). New York, NY, Oxford University Press. https://doi.org/10.1093/oso/9780190492908.003.0003

UNHCR. (2021). UNHCR statistical online population database. Data extracted: 12/02/2021. [https://www.unhcr.org/refugee-statistics/download/].

United Nations, Department of Economic and Social Affairs, Population Division. (2019a). World Population Prospects 2019, *Online Edition. Rev. 1.*

United Nations, Department of Economic and Social Affairs, Population Division. (2019b). World population prospects 2019: Methodology of the united nations population estimates and projections. https://population.un.org/wpp/assets/Files/WPP2019_Methodology.pdf

Unnevehr, L. J., & Stanford, M. L. (1985). Technology and the demand for women's labor in asian rice farming. In I. R. R. Institute (Ed.), *Women in rice farming: Proceedings of a conference* *on women in rice farming systems, the international rice research institute, p.o. box 933, manila, philippines, 26-30 september 1983* (pp. 1–18). Aldershot, Hants, England, Gower.

Vickery, M. (1983). Democratic kampuchea: Themes and variations. In D. P. Chandler & B. Kiernan (Eds.), *Revolution and its aftermath in kampuchea: Eight essays* (pp. 99–135). New Haven, Yale University Southeast Asia Studies.

Yi, J., Heckman, J. J., Zhang, J., & Conti, G. (2015). Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal*, *125*(588), F347–F371.

Zélity, B. (2023). Age diversity and aggregate productivity. *Journal of Population Economics*, *36*(3), 1863–1899.

# Appendix A

## Chapter 1 Appendix

This appendix provides detailed derivations, calibration procedures, and supplemental simulation results referenced in the main paper. Appendix A.1 presents the full derivations of the model equations used to characterize capital accumulation, output growth, and the evolution of population and skill composition over time. Appendix A.2 outlines the calibration steps and data sources used to construct the population time path and skill distribution. Appendix A.3 contains additional figures and tables for reference.

## A.1 Model derivations

### A.1.1 Derivation of initial capital stock

This section derives the expression for the initial capital stock based on the model's capital accumulation equation. The equation incorporates differentiated saving rates across skill types. In the standard Solow's framework, the capital stock in the next period evolve according

$$K_{t+5} \quad = \quad sY_t + (1-\delta)K_t, \tag{A.1.5}$$

where $s$ is the aggregate saving rate and $\delta$ is the depreciation rate. This formulation assumes a uniform saving rate applied to total output. In contrast, the setting in this paper considers that income is derived from labor (disaggregated by skill type), land, and physical capital. Each income component contributes to capital accumulation according to its own saving rate. Therefore, the law of motion for capital becomes

$$K_{t+5} \quad = \quad \sum_{i=15}^{65} \left(\sigma_i^H w_{i,t}^H P_{i,t}^H + \sigma_i^L w_{i,t}^L P_{i,t}^L\right) + \sigma^K r_t K_t + \sigma^X q_t X_t + (1-\delta)K_t, \tag{A.1.6}$$

where $\sigma^K$ and $\sigma^X$ are the saving rates from capital and land income, respectively. $\sigma_i^H$ and $\sigma_i^L$ represent the saving rates of high-skilled and low-skilled workers in cohort $i$, respectively.

Reorganizing the saving terms and incorporation the model's wage structure, the capital accumulation can be expressed as

$$K_{t+5} = \sum_{i=15}^{65} \Phi_{i,t}^{\rho-\eta} \widetilde{w}_{i,t} P_{i,t} \left[\sigma_i^H \lambda \theta_{i,t}^{\eta} + \sigma_i^L (1-\lambda)(1-\theta_{i,t})^{\eta}\right] + \sigma^K \alpha Y_t + \sigma^X (1-\alpha-\beta) Y_t + (1-\delta)K_t, \tag{A.1.7}$$

where $\Phi_{i,t} \in [0,1]$ and and $\Phi_{i,t} = (1-\lambda)(1-\theta_{i,t})^{\eta} + \lambda \theta_{i,t}^{\eta}$. It is a skill adjustment factor that scales the cohort size based on its internal skill composition.

Substituting the wage structure, Equation (A.1.7) simplifies to

$$K_{t+5} = Y_t L_t^{-\rho} \beta \sum_{i=15}^{65} \Phi_{i,t}^{\rho-\eta} P_{i,t}^{\rho} \left[\sigma_i^H \lambda \theta_{i,t}^{\eta} + \sigma_i^L (1-\lambda)(1-\theta_{i,t})^{\eta}\right] + \sigma^K \alpha Y_t + \sigma^X (1-\alpha-\beta) Y_t + (1-\delta) K_t. \tag{A.1.8}$$

Grouping all saving terms and applying the production structure leads to

$$K_{t+5} = K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta} \left[\beta \widetilde{\Omega}_t + \alpha \sigma^K + (1-\alpha-\beta)\sigma^X\right] + (1-\delta) K_t, \tag{A.1.9}$$

which can be written as

$$K_{t+5} = \widetilde{\sigma}_t K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta} + (1-\delta) K_t, \tag{A.1.10}$$

where

$$\widetilde{\sigma}_t = \beta \widetilde{\Omega}_t + \alpha \sigma^K + (1-\alpha-\beta)\sigma^X, \tag{A.1.11}$$

and

$$\widetilde{\Omega}_t = \frac{\sum_{i=15}^{65} \gamma_i \Phi_{i,t}^{\rho-\eta} \left[\sigma_i^H \lambda \theta_{i,t}^{\eta} + \sigma_i^L (1-\lambda)(1-\theta_{i,t})^{\eta}\right] P_{i,t}^{\rho}}{\sum_{i=15}^{65} \gamma_i \Phi_{i,t}^{\rho} P_{i,t}^{\rho}}. \tag{A.1.12}$$

To derive the initial capital stock, it is assumes that capital, effective labor, and land all growth at constant rates, denoted by $g_K$, $g_L$, and $g_X$, respectively. This implies the following growth dynamics

$$L_{t+5} = (1+g_L) L_t, \tag{A.1.13}$$

$$K_{t+5} = (1+g_K) K_t, \tag{A.1.14}$$

$$X_{t+5} = (1+g_X) X_t. \tag{A.1.15}$$

Rewriting the capital accumulation equation in terms of growth yields

$$\frac{K_{t+5}}{K_t} = \widetilde{\sigma}_t K_t^{\alpha-1} L_t^{\beta} X_t^{1-\alpha-\beta} + (1-\delta) = 1 + g_K. \tag{A.1.16}$$

The left side of Equation (A.1.16) implies that the output to capital ratio at time $t$ can be written as

$$K_t^{\alpha-1} L_t^{\beta} X_t^{1-\alpha-\beta} = \frac{g_K + \delta}{\widetilde{\sigma}_t}. \tag{A.1.17}$$

Similarly, at time $t+5$

$$K_{t+5}^{\alpha-1} L_{t+5}^{\beta} X_{t+5}^{1-\alpha-\beta} \quad = \quad \frac{g_K + \delta}{\widetilde{\sigma}_{t+5}}. \tag{A.1.18}$$

At the steady state, the output-to-capital ratio remains constant over time. Accordingly, the initial capital stock $K_0$ can be expressed as

$$K_0 \quad = \quad \left[ \left( \frac{\widetilde{\sigma}_0}{g_K + \delta} \right) L_0^{\beta} X_0^{1-\alpha-\beta} \right]^{\frac{1}{1-\alpha}}. \tag{A.1.19}$$

In the initial period, land input is normalized to $X_0 = 1$, which simplifies the expression without loss of generality. This expression illustrates how the initial capital stock responds to saving behavior and factor input levels. Higher saving rates increase the resources allocated to investment, while a younger or more skilled population raises effective labor input. To maintain balanced growth and avoid capital scarcity, the model requires a larger capital stock to complement the higher productivity of labor at the start of the transition.

### A.1.2 Growth rate of capital

The growth rate of effective labor $g_L$ can be calculated directly from demographic data. Given this, the steady state value of $g_K$ can be derived by taking the ratio of Equation (A.1.18) to Equation (A.1.17). Applying the growth assumptions in Equations (A.1.13)-(A.1.15) yields

$$1 + g_K \quad = \quad \left[ (1 + g_{\widetilde{\sigma}}) (1 + g_L)^{\beta} (1 + g_X)^{1-\alpha-\beta} \right]^{\frac{1}{1-\alpha}}, \tag{A.1.20}$$

where $1 + g_{\widetilde{\sigma}} = \frac{\widetilde{\sigma}_{t+5}}{\widetilde{\sigma}_t}$ represents the gross growth rate of total saving in the economy.

Under the assumption that the economy was on a balanced growth path in 1950, the effective saving rate is constant over time. This implies $g_{\widetilde{\sigma}} = 0$, and $1 + g_{\widetilde{\sigma}} = 1$. Substituting this condition simplifies the expression to

$$1 + g_K \quad = \quad \left[ (1 + g_L)^{\beta} (1 + g_X)^{1-\alpha-\beta} \right]^{\frac{1}{1-\alpha}}. \tag{A.1.21}$$

The Equation (A.1.21) shows that capital must grow in proportion to labor and land in order to maintain constant factor shares. Faster growth in effective labor increases the need for capital accumulation to prevent capital from becoming relatively scarce. If the labor force becomes more productive or grows larger, investment must rise accordingly to preserve a stable output path.

### A.1.3 Growth rate of GDP per capita

Output per capita at time $t$ is defined as

$$y_t = \frac{K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta}}{P_t}, \tag{A.1.22}$$

and output per capital at time $t+5$ is given by

$$y_{t+5} = \frac{K_{t+5}^{\alpha} L_{t+5}^{\beta} X_{t+5}^{1-\alpha-\beta}}{P_{t+5}}. \tag{A.1.23}$$

Taking ratio of these two expressions yields

$$\frac{y_{t+5}}{y_t} = \left(\frac{K_{t+5}}{K_t}\right)^{\alpha} \left(\frac{L_{t+5}}{L_t}\right)^{\beta} \left(\frac{X_{t+5}}{X_t}\right)^{1-\alpha-\beta} \left(\frac{P_t}{P_{t+5}}\right). \tag{A.1.24}$$

Rewrite in terms of gross growth rates gives

$$1 + g_y = (1+g_K)^{\alpha} (1+g_L)^{\beta} (1+g_X)^{1-\alpha-\beta} \left(\frac{1}{1+g_P}\right), \tag{A.1.25}$$

where $g_P$ is the population growth rate, and $g_y$ represents the net growth rate of GDP per capita.

Equation (A.1.25) shows that GDP per capita increases when the growth of capital and effective labor exceeds population growth, assuming land grows at a constant rate. This condition ensures that the gains from factor accumulation are not offset by the rising number of individuals sharing the output. When capital and labor expand more rapidly than the population, income per capita rises. Conversely, if factor accumulation is too slow relative to demographic pressures, living standards stagnate or decline.

### A.1.4 Growth rate of land productivity

In the model, the economy is assumed to be on a balanced growth path prior to the genocide. To ensure that all production factors contribute proportionally to long-run output growth, the growth rate of land productivity $g_X$ must be consistent with this assumption.

The steady state value of $g_X$ is derived by combining the expressions for capital accumulation and GDP per capita growth. Substituting Equation (A.1.21) into Equation (A.1.25) and rearrange terms yields

$$1 + g_X = \left[(1+g_y)^{1-\alpha} (1+g_P)^{1-\alpha} (1+g_L)^{-\beta}\right]^{\frac{1}{1-\alpha-\beta}}. \tag{A.1.26}$$

This expression determines how fast land productivity must grow beginning in the next period to preserve constant factor shares. While the level of land is fixed at 1 in the initial period, its subsequent productivity growth must keep pace with other productive factors. If output per capita and population grow rapidly, land must expand accordingly to avoid scarcity, maintain production efficiency, and sustain balanced growth over time.

### A.1.5 Age-specific migration factor

Age-specific migration factors, $m_{i,t}$, are not directly observed from the data. Instead, they are calculated as residuals from the dynamic age distribution equation.

Recall Equation (1.11) in the main paper, which states that the age-5 cohort in period $t+5$ is $P_{0,t+5} = s_{\text{1-4},t} m_{0,t} P_{0,t}$. Rearranging this expression, the migration factor for the newborn cohort is computed as

$$m_{0,t} = \frac{P_{5,t+5}}{s_{\text{0-1},t} P_{0,t}}, \tag{A.1.27}$$

where $s_{\text{0-1},t}$ is the survival rate from ages 0 to 1. As discussed in the main text, the survival rate for the newborn cohort is adjusted to avoid over-correcting for infant mortality.

Similarly, for all other cohort $i$, the migration factor is given by

$$m_{i,t} \quad = \quad \frac{P_{i+5,t+5}}{s_{i,t} P_{i,t}}, \tag{A.1.28}$$

which corresponds to Equation (1.13) in the main paper, derived from the cohort transition equation $P_{i+5,t+5} = s_{i,t} m_{i,t} P_{i,t}$.

The aggregate net migration across all cohorts is computed as

$$\sum_{i=0}^{95} (m_{i,t} - 1) P_{i,t}. \tag{A.1.29}$$

A value of $m_{i,t} > 1$indicates net migration for cohort $i$, while $m_{i,t} < 1$ reflects net emigration.

This structure captures migration as the residual component of cohort population dynamics after accounting for survival. The model treats migration as an unobserved adjustment factor that adjusts simulated cohort sizes to fit observed population data. This approach enables flexible demographic transitions without requiring detailed migration records.

### A.1.6 Survival rate by skill type

The model allows survival rates to vary by skill type that reflecting demographic differences in mortality. The high-skilled survival rate for cohort $i$ in period $t$ is guessed by

$$s_{i,t}^{H} \quad = \quad (1 + \psi_{i,t})\, s_{i,t}, \tag{A.1.30}$$

where $\psi_{i,t} \in \left[-1, \frac{1}{s_{i,t}} - 1\right]$ is an exogenous parameter that capture a one-time mortality shock specific to the cohort $i$.

The aggregate survival rate for cohort $i$ is defined as a weighted average of the skill-specific survival rates

$$s_{i,t} \quad = \quad \theta_{i,t} s_{i,t}^{H} + (1 - \theta_{i,t})\, s_{i,t}^{L}, \tag{A.1.31}$$

where $\theta_{i,t} = \frac{P_{i,t}^{H}}{P_{i,t}}$ denotes the high-skilled share of the cohort.

Substituting the expression for $s_{i,t}^{H}$ and solving for the low-skilled survival rate yields

$$s_{i,t}^{L} \quad = \quad \frac{1 - (1 + \psi_{i,t})\, \theta_{i,t}}{1 - \theta_{i,t}} s_{i,t}. \tag{A.1.32}$$

The same structure is applied to the newborn cohort ($i = 0$) for the calculating newborn survival rate $s_{\text{0-1},t}$. The total number of newborns at time $t + 5$ is given by

$$P_{0,t+5} \quad = \quad P_{0,t+5}^{H} + P_{0,t+5}^{L}, \tag{A.1.33}$$

where

$$P_{0,t+5}^{H} = 5 \left( \sum_{i=15}^{45} s_{\text{0-1},t}^{H} n_{i,t} \pi^{H} m_{i,t} f_{i,t} P_{i,t}^{H} + \sum_{i=15}^{45} s_{\text{0-1},t}^{L} n_{i,t} \pi^{L} m_{i,t} f_{i,t} P_{i,t}^{L} \right), \tag{A.1.34}$$

and

$$P_{0,t+5}^{L} = 5 \left( \sum_{i=15}^{45} s_{\text{0-1},t}^{H} n_{i,t} (1 - \pi^{H}) m_{i,t} f_{i,t} P_{i,t}^{H} + \sum_{i=15}^{45} s_{\text{0-1},t}^{L} n_{i,t} (1 - \pi^{L}) m_{i,t} f_{i,t} P_{i,t}^{L} \right), \tag{A.1.35}$$

Grouping term gives, the total number of newborns becomes

$$P_{0,t+5} = 5 s_{\text{0-1},t}^{H} \left( \sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}^{H} \right) + 5 s_{\text{0-1},t}^{L} \left( \sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}^{L} \right). \tag{A.1.36}$$

Alternatively, this expression can be written using the aggregate survival rate $s_{\text{0-1},t}$

$$P_{0,t+5} \quad = \quad 5 \left( \sum_{i=15}^{45} s_{\text{0-1},t} n_{i,t} m_{i,t} f_{i,t} P_{i,t} \right), \tag{A.1.37}$$

where $P_{i,t} = P_{i,t}^{H} + P_{i,t}^{L}$. Equating the two expression and simplifying gives

$$s_{\text{0-1},t} \quad = \quad \tau s_{0,t}^{H} + (1 - \tau)\, s_{0,t}^{L}, \tag{A.1.38}$$

where

$$\tau_t = \frac{\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}^{H}}{\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}}, \tag{A.1.39}$$

denote the fraction of births from high-skilled parents. The complement, $1-\tau_t = \frac{\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}^L}{\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}}$ represents the share of births from low-skilled parents. Note that $\tau_t$ reflects the parental composition at birth, not the realized skilled type of the newborns themselves.

Because direct data on high-skilled newborn survival are unavailable, the model also assumes

$$s_{0\text{-}1,t}^H = (1+\psi_{0,t})\, s_{0\text{-}1,t}, \tag{A.1.40}$$

where $\psi_{0,t} \in \left[-1, \frac{1}{s_{0\text{-}1,t}} - 1\right]$ captures a one-time mortality shock. Substituting this into the expression above and solving for the low-skilled newborn survival rate yields

$$s_{0\text{-}1,t}^L = \frac{1-(1+\psi_{0,t})\,\tau_t}{1-\tau_t} s_{0\text{-}1,t}. \tag{A.1.41}$$

This structure allows survival rates to differ by parental skill type while ensuring consistency with aggregate demographic data. High-skilled survival rate is set as a proportional to the overall survival rate, and the low-skilled rate adjusts to preserve internal consistency.

The shock parameter $\psi_{i,t}$ captures mortality disruption affecting specific skill groups. Under normal conditions, the model assumes no differential mortality across skill types. In this scenario, setting $\psi_{i,t} = 0$ implies that $s_{i,t}^H = s_{i,t}^L = s_{i,t}$. This assumption reflects the pre-genocide period when mortality risks were not systematically associated with education. As a result, the share of high-skilled $\theta_{i,t}$ remains constant over time, since skill-specific survival rates do not alter cohort composition in subsequent periods.

In 1975, however, the Khmer Rouge regime specifically targeted educated population, disproportionately affecting high-skilled individuals. To capture this historical event, the model introduces a one-time negative survival shock, $\psi_{i,t} < 0$, for affected cohorts. Specifically, it sets

$$\psi_{i,t} = \begin{cases} -0.25 & \text{if} \quad i \in \{15, 20, ..., 95\}, t = 1975, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1.42}$$

This shock corresponds to a 25% reduction in the survival rate of high-skilled individuals compared to low-skilled individuals in 1975. Thus, the relationship between survival rates across skill types becomes

$$s_{i,t}^H \begin{cases} < \quad s_{i,t}^L & \text{if} \quad i \in \{15, 20, ..., 95\}, t = 1975 \\ = \quad s_{i,t}^L & \text{otherwise.} \end{cases} \tag{A.1.43}$$

The 25% mortality penalty is motivated by historical evidence discussed in the main text.

### A.1.7 Share of high-skilled newborn

This section derives the fraction of newborns who are high-skilled, denoted by $\theta_{0,t+5}$. The high-skilled share of cohort $i$ in period $t$ is defined as

$$\theta_{i,t} = \frac{P^H_{i,t}}{P_{i,t}}. \tag{A.1.44}$$

Using the dynamic age equation from above, the high-skilled share of cohort $i+5$ at time $t+5$ evolves according to

$$\theta_{i+5,t+5} = \frac{s^H_{i,t} m_{i,t} P^H_{i,t}}{s_{i,t} m_{i,t} P_{i,t}} = \left(\frac{s^H_{i,t}}{s_{i,t}}\right) \theta_{i,t}. \tag{A.1.45}$$

Substituting the expression for $s^H_{i,t}$ from Equation (A.1.30), this relationship simplifies to

$$\theta_{i+5,t+5} = (1 + \psi_{i,t})\, \theta_{i,t}. \tag{A.1.46}$$

The share of high-skilled newborns at time $t+5$, denoted by $\theta_{0,t+5}$, is defined as

$$\theta_{0,t+5} = \frac{P^H_{0,t+5}}{P_{0,t+5}}. \tag{A.1.47}$$

Expanding the numerator and denominator using Equations (A.1.34) and (A.1.37), the expression becomes

$$\theta_{0,t+5} = \frac{5 s^H_{0\text{-}1,t} \left(\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P^H_{i,t}\right) + 5 s^L_{0\text{-}1,t} \left(\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P^L_{i,t}\right)}{5 s_{0\text{-}1,t} \left(\sum_{i=15}^{45} n_{i,t} m_{i,t} f_{i,t} P_{i,t}\right)}. \tag{A.1.48}$$

Substitution the expressions for $s^H_{0\text{-}1,t}$ and $s_{0\text{-}1,t}$ from Equations (A.1.40) and (A.1.38), and simplifying yields

$$\theta_{0,t+5} = (1 + \psi_{0,t})\, \tau_t \pi^H + \left[1 - (1 + \psi_{0,t})\, \tau_t\right] \pi^L. \tag{A.1.49}$$

Even if high-skilled parents represent a large share of the reproductive population or have higher fertility, a negative shock to their survival rates reduces the share of high-skilled newborns. The equation captures how selective mortality during periods of conflict or crisis can offset intergenerational human capital transmission and lower the skill composition of future cohorts, even when the fertility behavior of high-skilled parents would otherwise increase their population share.

### A.1.8 Derivation of $\theta_0$

In the model, the economy is assumed to be on a balanced growth path, which implies that in the baseline year 1950, there are no population shocks. Consequently, the shock parameter is set to zero, $\psi_{i,0} = 0$. Under this condition, the skill share is constant across cohorts in 1950. Therefore, $\theta_{i,0} = \theta_0$, indicating that the share of high-skilled individuals is identical across all cohorts. As a result, all cohort-level characteristics remain constant in steady state, which implies $\theta_{0,t+5} = \theta_{0,t} = \theta_0$. Accordingly, Equation (A.1.49) becomes

$$\theta_0 = \tau_0 \pi^H + (1 - \tau_0)\,\pi^L. \tag{A.1.50}$$

Because the economy is on its balanced growth path, the skill share $\theta_{i,t}$ is constant across both cohort and time. Hence, Equation (A.1.39) also simplifies to

$$\tau_0 = \frac{\sum_{i=15}^{45} n_{i,0} m_{i,0} f_{i,0} P_{i,0}^H}{\sum_{i=15}^{45} n_{i,0} m_{i,0} f_{i,0} P_{i,0}}. \tag{A.1.51}$$

By the definition, the number of high-skilled individual in cohort $i$ at steady state is given by $P_{i,0}^H = \theta_{i,0} P_{i,0}$. Substitute this into Equation (A.1.51) gives

$$\tau_0 = \frac{\sum_{i=15}^{45} n_{i,0} m_{i,0} f_{i,0} \theta_0 P_{i,0}}{\sum_{i=15}^{45} n_{i,0} m_{i,0} f_{i,0} P_{i,0}} = \theta_0. \tag{A.1.52}$$

Substitute $\tau_0 = \theta_0$ into Equation (A.1.50) yields

$$\theta_0 = \theta_0 \pi^H + (1 - \theta_0)\,\pi^L. \tag{A.1.53}$$

This can be solved for $\theta_0$ as

$$\theta_0 = \frac{\pi^L}{1 - (\pi^H - \pi^L)}. \tag{A.1.54}$$

This expression characterizes the steady-state share of high-skilled individuals in the population, determined solely by the intergenerational transmission parameters $\pi^H$ and $\pi^L$. At the same time, on the balanced growth path, the skill composition of each cohort $\theta_0$ remains constant over time. In the absence of differential survival rates across skill types, high-skilled and low-skilled individuals face identical mortality risks, and thus their relative shares evolve uniformly across generations.

### A.1.9 Derivation of special case I: Without skill heterogeneity

To isolate the contribution of skill composition to long-run economic outcomes, this section considers a special case where skill heterogeneity is removed. Specifically, the productivity parameter $\lambda = 0.5$. This means that high-skilled and low-skilled labor are equally productive. As a result, the labor input becomes homogeneous, while all other parameters and model assumptions remain unchanged from the baseline specification.

Under this assumption, the production function is

$$Y_t = K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta}. \tag{A.1.55}$$

The labor input is defined as

$$L_t = \left[\sum_{i=0}^{95} \gamma_i P_{i,t}^{\rho}\right]^{\frac{1}{\rho}}. \tag{A.1.56}$$

The corresponding factor prices are

$$w_{i,t} = \beta\gamma_i \frac{Y_t}{L_t}\left(\frac{L_t}{P_{i,t}}\right)^{1-\rho}, \tag{A.1.57}$$

$$r_t = \alpha \frac{Y_t}{K_t}, \tag{A.1.58}$$

$$q_t = (1-\alpha-\beta)\frac{Y_t}{X_t}. \tag{A.1.59}$$

Capital accumulation follows

$$K_{t+5} = \sum_{i=15}^{65} \sigma_i^w w_{i,t} P_{i,t} + \sigma^K r_t K_t + \sigma^X q_t X_t + (1-\delta)K_t, \tag{A.1.60}$$

where $\sigma_i^w$ is the saving rates of labor incomes for cohort $i$. $\sigma^K$ and $\sigma^X$ represent the saving rates of capital and land incomes, respectively.

Substituting the expressions for the factor prices from Equations (A.1.57), (A.1.58), and (A.1.59), the capital accumulation equation becomes

$$K_{t+5} = Y_t \beta \frac{1}{L_t^{\rho}} \sum_{i=15}^{65} \sigma_i^w \gamma_i P_{i,t}^{\rho} + \sigma^K \alpha Y_t + \sigma^X (1-\alpha-\beta) Y_t + (1-\delta)K_t. \tag{A.1.61}$$

Using the production function in Equation (A.1.55), this can be written as

$$K_{t+5} = \sigma_t K_t^{\alpha} L_t^{\beta} X_t^{1-\alpha-\beta} + (1-\delta)K_t, \tag{A.1.62}$$

where

$$\sigma_t = \beta\Omega_t + \alpha\sigma^K + (1-\alpha-\beta)\sigma^X, \tag{A.1.63}$$

and

$$\Omega_t = \frac{\sum_{i=15}^{65} \sigma_i^w \gamma_i P_{i,t}^{\rho}}{\sum_{i=15}^{65} \gamma_i P_{i,t}^{\rho}}. \tag{A.1.64}$$

The term $\sigma_t$ represents the saving rate from all factor incomes in the production function.

Given that capital, labor, and land grow at constant rates, the initial capital stock and the its growth rates can be derived as

$$K_0 = \left[\left(\frac{\sigma_0}{g_K+\delta}\right) L_0^{\beta} X_0^{1-\alpha-\beta}\right]^{\frac{1}{1-\alpha}}, \tag{A.1.65}$$

and

$$1+g_K = \left[(1+g_\sigma)(1+g_L)^{\beta}(1+g_X)^{1-\alpha-\beta}\right]^{\frac{1}{1-\alpha}}. \tag{A.1.66}$$

Under balance growth path assumption, saving is constant, implying $g_\sigma = 0$, and hence $1+g_\sigma = 1$. Substituting this into the equation simplifies the growth rate of capital to

$$1+g_K = \left[(1+g_L)^{\beta}(1+g_X)^{1-\alpha-\beta}\right]^{\frac{1}{1-\alpha}}. \tag{A.1.67}$$

The growth rate of output per capita is given by

$$\frac{y_{t+5}}{y_t} = \left(\frac{K_{t+5}}{K_t}\right)^{\alpha}\left(\frac{L_{t+5}}{L_t}\right)^{\beta}\left(\frac{X_{t+5}}{X_t}\right)^{1-\alpha-\beta}\left(\frac{P_t}{P_{t+5}}\right). \tag{A.1.68}$$

In terms of growth rates, this becomes

$$1+g_y = (1+g_K)^{\alpha}(1+g_L)^{\beta}(1+g_X)^{1-\alpha-\beta}\left(\frac{1}{1+g_P}\right). \tag{A.1.69}$$

Rearranging Equation (65) to solve for the growth rate of land productivity gives

$$1+g_X = \left[(1+g_y)^{1-\alpha}(1+g_P)^{1-\alpha}(1+g_L)^{-\beta}\right]^{\frac{1}{1-\alpha-\beta}}. \tag{A.1.70}$$

### A.1.10 Derivation of special case II: Without skill heterogeneity and capital

This section considers a second special case in which both capital and skill heterogeneity are removed from the model. Specifically, the productivity parameter is set $\lambda = 0.5$, and the capital share is set to $\alpha = 0$. As a result, the production function simplifies to a Malthusian economy with only labor and land as inputs. There is no capital accumulation, and the economy operates without savings. Workers receive wages and consume their entire income.

The production function in this setting is

$$Y_t = L_t^{\beta} X_t^{1-\beta}, \tag{A.1.71}$$

where $\beta = 0.9$ is the labor share, and $1-\beta = 0.1$ is the land share. The initial land input $X_t$ is normalized to 1 consistent with the baseline model. All other parameter values remain the same unless otherwise noted.

The effective labor input is defined as

$$L_t = \left[\sum_{i=0}^{95} \gamma_i P_{i,t}^{\rho}\right]^{\frac{1}{\rho}}. \tag{A.1.72}$$

The factor prices in this setting are given by

$$w_{i,t} = \beta\gamma_i \frac{Y_t}{L_t}\left(\frac{L_t}{P_{i,t}}\right)^{1-\rho}, \tag{A.1.73}$$

$$q_t = (1-\beta)\frac{Y_t}{X_t}. \tag{A.1.74}$$

Because capital is not part of the economy, output per capita is determined entirely by the dynamics of labor and land. The growth rate of output per capita is expressed as

$$1+g_y = (1+g_L)^{\beta}(1+g_X)^{1-\beta}\left(\frac{1}{1+g_P}\right). \tag{A.1.75}$$

This expression implies that output per capita increases when the growth rates of effective labor and land productivity exceeds population growth.

Rearranging Equation (A.1.75) to solve for the growth rate of land productivity yields

$$1+g_X = \left[\frac{(1+g_y)(1+g_P)}{(1+g_L)^{\beta}}\right]^{\frac{1}{1-\beta}}. \tag{A.1.76}$$

This expression captures the rate at which land productivity must grow to maintain balanced growth in an economy without capital accumulation or skill differentiation. When labor becomes more productive or grows rapidly, land productivity must grow correspondingly to avoid diminishing returns and sustain improvements in per capita income.

## A.2 Data and calibration

### A.2.1 Calibrating $\gamma_i$

The productivity weights $\gamma_i$ are calibrated by solving for the set of age-specific efficiencies that best match model-implied labor productivity across cohorts. The goal is to replicate the life-cycle productivity profile observed in the data.

The calibration procedure begins with a uniform initial guess, setting each $\gamma_i$ to 0.1 for working-age cohorts. The system of nonlinear equations is then solved using the Levenberg–Marquardt algorithm, implemented via MATLAB's `fsolve` function. The objective is to align simulated cohort productivity with observed demographic and wage data.

Once the optimal values are obtained, $\gamma_i$ is extended to cover all age groups by assigning zero productivity to individuals under age 15 and over age 69. The weights are then normalized so that their sum across working-age cohorts satisfies

$$\sum_{i=15}^{65} \gamma_i = 1. \tag{A.2.1}$$

This calibration ensures that labor input in the model captures age-specific variations in productivity, consistent with the empirically observed concave age–earnings profile.

### A.2.2 Calibrating $\lambda$

The 2010 Cambodia Socio-Economic Survey (CSES), obtained from Humphreys (2015), contains a total of 3,311 observation. The sample ranges from ages 8 to 79. Among them, 3,251 individuals fall within the working-age range of 15 and 69. Of this group, 422 individuals have completed more than 12 years of schooling and are classified as high-skilled. The wage variable is measured as the hourly wage in U.S. dollars.

To estimate the earnings profile, a regression with robust standard errors is applied using the following specification

$$\log w_i \quad = \quad \beta_1 \text{age}_i + \beta_2 \ \text{age}_i^2 + \beta_3 \ \text{skill}_i + \beta_4 \ \text{male}_i + \mu_i, \tag{A.2.2}$$

where $\text{age}_i$ is individual's age. $\text{skill}_i$ is a dummy equal to 1 for high-skilled workers (those with more than 12 years of schooling). $\text{male}_i$ is a dummy for male, and $\mu_i$ is the error term.

Predicted log wages are transformed back into hourly wage units. These are then averaged by skill and age groups. Figures A.11 and A.12 present the resulting wage profiles for high-skilled and low-skilled individuals, and the overall average wage profile.

Calibration of the relative productivity parameter $\lambda$ proceeds by normalizing predicted wages in each age group relative to the wage of 15-19 cohort. The relative productivity weight for cohort $j$ is defined as

$$\widehat{\lambda}_j \quad = \quad \frac{\widehat{w}_j^H}{\widehat{w}_j^H + \widehat{w}_{j,t}^L \left(\frac{\widehat{\theta}_j}{1-\widehat{\theta}_j}\right)^{\eta-1}}, \tag{A.2.3}$$

where $j \in \{15, \ldots, 65\}$, $\widehat{w}_j^H$ and $\widehat{w}_j^L$ are predicted wages for high-skilled and low-skilled individuals, respectively. $\widehat{\theta}_j$ is the predicted fraction of high-skilled workers in cohort $j$. The substitution parameter $\eta$ is set to 0.9 following standard assumption in the model.

The final value for $\lambda$ is obtained by average $\widehat{\lambda}_j$ across all cohorts $j$. The average predicted value is approximately $\widehat{\lambda} = 0.67$.

### A.2.3 High-skilled and low-skilled wages

High-skilled and low-skilled wage profiles are constructed based on the predicted hourly wages from the wage regression described above in Section A.2.2. Individuals are grouped into five-year age categories: 15-19, 20-24, ..., 65-69. An individual is classified as high-skilled if they have completed more than 12 years of schooling. Average predicted wages are computed separately for each skill group within each age bracket.

The composite wage profile for each cohort $j \in \{15, \ldots, 65\}$ in the 2010 CSES data is constructed as

$$\bar{w}_j = \left(1 - \widehat{\theta}_j\right) \widehat{w}_j^L + \widehat{\theta}_j \widehat{w}_j^H, \tag{A.2.4}$$

where $\widehat{w}_j^L$ and $\widehat{w}_j^H$ are the predicted low-skilled and high-skilled wages, respectively. $\widehat{\theta}_j$ is the predicted share of high-skilled individuals in cohort $j$.

These composition wage profiles are used to calibrate the age-specific productivity weights $\gamma_i$. Wages implied by the production function depend on aggregate effective labor, which incorporates age-specific productivity weights. Calibration proceed by choosing $\gamma_i$ to minimize the sum of squared deviations between observed cohort-level wages $\bar{w}_j$ and those implied by the production function.

This approach ensures internal consistency between the wage profiles observed in the data and those generated from production function.

### A.2.4 Saving rate

This section constructs saving rate profiles based on cohort effects estimated by Deaton and Paxson (2000), due to the lack of age-specific saving rate data for Cambodia. The authors use data from Thailand's Socio-Economic Surveys conducted in 1976, 1981, 1986, 1988, 1990, and 1992 to study the evolution of saving behavior over the life cycle. Their findings show that saving increases with age and differs systematically across cohorts, producing a hump-shaped pattern consistent with the lifecycle hypothesis.

Figure 2 of Deaton and Paxson (2000) reports predicted cohort effects on $\ln y - \ln c$, based on their estimated model (Equation 14). These log differences serve as a proxy for saving intensity, capturing the deviation of consumption from income in logarithmic terms. The approximate log saving intensities for cohorts aged 5 and 65 are $-0.12$ and 0.18, respectively. A smooth saving intensity profile $\kappa_i$ is constructed by linear interpolation between these two endpoints (see Figure A.13).

To convert $\kappa_i$ into a saving rate bounded between zero and one, I apply the exponential transformation

$$\sigma_i = 1 - \exp^{(-\kappa_i)}, \tag{A.2.5}$$

where $i \in [5, ..., 65]$ and $\sigma_i$ denotes the average saving rate for cohort $i$ and $\kappa_i \in [-0.02, ..., 0.18]$.

The model incorporates heterogeneity in saving behavior by skill type. Let $\sigma_i^L$ and $\sigma_i^H$ are the saving rates of low-skilled and high-skilled individuals, respectively, for cohort $i \in \{15, \ldots, 65\}$. To decompose skill-specific saving rates, I assume that high-skilled individuals save twice as much as low-skilled individuals

$$\sigma_i^H \quad = \quad 2\sigma_i^L. \tag{A.2.6}$$

I further impose that the cohort-level average saving rate $\sigma_i$ equals the mean of the two group-specific saving rates

$$\sigma_i \quad = \quad \frac{1}{2}\left(\sigma_i^H + \sigma_i^L\right). \tag{A.2.7}$$

Substituting the first condition into the second yields

$$\sigma_i^L \quad = \quad \frac{2}{3}\sigma_i. \tag{A.2.8}$$

Figure A.14 illustrates the reconstructed saving rate profile for low-skilled, high-skilled, and average individuals across age cohorts.

The assumption that high-skilled individuals save twice as much as low-skilled individuals is motivated by human capital theory and by the observation from countries without universal health care or social insurance. In such contexts, higher-income individuals tend to save larger portion of their earnings. This provides a rationale for assigning higher saving rates to high-skilled individuals. In Cambodia, as in neighboring Thailand, households often rely on personal savings to finance essential expenditures such as health care, education, and family support. Given these structural features, it is reasonable to assume that high-skilled individuals not only earn more, but also save more than their low-skilled counterparts.

### A.2.5 Age-specific fertility rate

Age-specific fertility data are available for five-year age groups of women between ages 15 and 49, reported at five-year intervals from 1950 to 2020. Each observation represents the number of births per $1,000$ women in a given age group during the corresponding five-year interval (see Section 3 of main paper for detail).

To isolate long-run demographic trends from the impact of the Cambodian genocide and its aftermath, the fertility rates for the years 1970 to 1990 are treated as missing. This adjustment removes the effects of both the mortality shock and the post-genocide replacement fertility observed during that period.

The missing values are then interpolated using a cubic spline method to construct a smoothed counterfactual fertility series. This approach produces a gradual and continuous evolution in fertility behavior over time. This is consistent with the assumption that, in the absent of genocide, fertility trends would have exhibited a smoother long-run path.

### A.2.6 Survival rate

Age-specific survival probabilities are sourced directly from the Abridged Life Tables (ALT). These tables report survival rates across standard age intervals, starting with ages 0-1 and 1-4, followed by five-year cohorts (e.g., 5-9, 10-14, . . ., 95-99) as described in Section 3 of the main paper. Since neither the population data nor the model distinguish by sex, the same age-specific survival rates are applied to males and females within each cohort.

To simulate demographic dynamics in the absence of the Cambodian genocide, survival rates for the years 1970, 1975, and 1980 are treated as missing. This adjustment removes the sharp mortality shock associated with the Khmer Rouge regime. Moreover, the missing values are interpolated using a cubic spline method based on survival rates from the adjacent years. This approach produces a smooth and continuous survival rate trajectory. It is based on the premise that, in the absence of the genocide, survival rates would have followed a gradual upward trend rather than exhibiting a sudden decline. These interpolated trajectories provide serve as a counterfactual baseline for estimating what population dynamics might have looked like in the absence of the mortality shock.

To construct counterfactual survival rates by skill type, I substitute the counterfactual overall survival rate into Equations (A.1.30), (A.1.32), (A.1.40), and (A.1.41), substituting in the counterfactual overall survival rate calibrated from the interpolated values. This ensures that skill-specific survival paths are consistent with the counterfactual demographic scenario.

### A.2.7 Age-specific migration factor

Since the data for age-specific migration data is not available, migration factors are calibrated using Equations (A.1.27) and (A.1.28) in Section A.1.5. Once age-specific migration factors are simulated, aggregate migration by year is constructed using Equation (A.1.29). This enables an external validity check by comparing with net migration estimates from the United Nations (UN).[9]

Figure A.1 presents a comparison between the constructed net migration series (green dashed line) and the UN estimates (blue solid line). The two series display broadly similar patterns, although the constructed series slightly overestimates emigration during the early 1980s. However, determining which estimate is more accurate is not straightforward, as each relies on distinct data sources and estimation methodologies.

According to the United Nations High Commissioner for Refugee (UNHCR, 2021), the number of Cambodian refugees between 1980 and 1984 was approximately $615,000$, nearly double the $330,000$ refugees reported between 1975 and 1979. This suggests that actual emigration during 1980-1984 may have been higher than what is reflected in the UN population dataset, which may understate the magnitude of refugee flows during that period.

There are several plausible explanations for the discrepancies between the constructed and UN migration series. First, neither dataset relies on direct observations of inflows and outflows. Instead, both series use population accounting identities to infer net migration as a residual. According to the

[9]The United Nations aggregate net migrants data is extracted from: https://population.un.org/wpp.

methodology of the United Nations Population Division (2019b), net migration is estimated as the residual difference between census rounds, adjusted for natural increase.[10] Since both approaches rely on residual estimation, the constructed series is not expected to perfectly match the UN data.

Second, inaccuracies in mortality rates, particularly during and after the genocide, can distort inferred migration. Because migration is inferred as the unexplained component in the population dynamics, even small inaccuracies in survival probabilities may significantly bias the results. Third, births during the Khmer Rouge regime may have been underreported. Such underreporting would reduce estimated population sizes and thus inflate the inferred level of net migration.

The reconstructed age-specific migration factors display volatility across cohorts, as shown in Figure A.2. This variation arises because migration is calculated as a residual from observed cohort sizes and survival probabilities, rather than being directly observed.

Although age-specific migration does not affect the observed population data directly, it plays a key role in shaping the counterfactual demographic scenario. Because the counterfactual excludes mass killings during the Pol Pot regime, there is no empirical basis for what migration would have looked like during that period. As a result, instead of using interpolation, the counterfactual age-specific migration factors are assumed to equal the average values over the 1950-2015 period (see Figure A.2). This approach helps eliminate the extreme emigration observed during the genocide and the vast immigration that occurred in the post-genocide period, particularly in the 1990s.

### A.2.8 Calibrating $\theta_{i,t}$

The initial value of $\theta_{i,t}$ is calibrated using Equation (A.1.54), which provides a closed form steady state solution. Subsequent value of $\theta_{i,t}$ evolve recursively using Equation (A.1.46), which depends on shock parameter $\psi_{i,t}$ and the preceding values of $\theta_{i,t}$.

To calibrate $\theta_0$, the parameters $\pi^H = 0.3$ and $\pi^L = 0.1$ are chosen based on the assumptions and justification provided in Section 5.3 of the main paper. These values ensure that the calibrated value of $\theta_0$ matches approximately 12.42%.

This benchmark is derived from the 2010 CSES, using both education and wage data. In the sample, approximately 12.42% of individuals report more than 12 years of schooling, which defines the high-skilled group. To validate this estimate, the share of high-skilled workers is also inferred from observed average wages using the wage identity introduced in Section A.2.3.

### A.2.9 Population time path

#### A.2.9.1 Actual population

To reconstruct the actual population time path, the simulation begins with the initial population in 1950, structured in five-year age groups (0-4, 5-9, 10-14, ..., 95-99), and sourced from the data. The population then evolves forward in five-year intervals using the population dynamics Equation

[10] See UN WPP 2019 Methodology for further detail https://population.un.org/wpp/assets/Files/WPP2019_Methodology.pdf.

(1.13) of the main paper. The fraction of females in each cohort is also calculated from the data and used in fertility computations.

The simulation proceeds using Equations from the main paper as follows:

- For each five-year period, beginning with the initial 1950 population, Equation (1.11) is used to compute the number of newborns in the next period (e.g., 1955) by aggregating fertility contributions from all reproductive-age cohorts.

- Newborns remain in the 0-4 age group during the period in which they are born and transition to the 5-9 group in the following period, as specified by Equation (1.12). All other cohorts are similarly aged forward by one group every five years, subject to adjustments for survival and migration according to Equation (1.13).

- This process is applied recursively. Simulated population values from each period become the inputs for calculating cohort sizes in the next. The simulation continues in five-year steps until the population is projected through the year 2015.

#### A.2.9.2 Counterfactual population

The procedure for simulating the counterfactual population time path mirrors that of the actual population simulation but incorporates key adjustments to remove the demographic impacts of the Khmer Rouge regime. The objective is to estimate how the population would have evolved in the absence of the genocide.

To this end, the counterfactual simulation uses fertility, survival, and migration rates calibrated under a no-genocide scenario.

The construction of these counterfactual inputs is detailed in Section A.2.5 for fertility, Section A.2.6 for survival rates, and Section A.2.7 for migration factors.

#### A.2.9.3 Population by skill type

The simulation of population by skill type builds on the baseline population dynamics outlined earlier. It distinguishes individuals into high-skilled and low-skilled groups based on their skill-specific survival rates and intergenerational transmission dynamics.

In 1950, the high-skilled and low-skilled populations are initialized by multiplying the total population in each cohort by the initial high-skilled share $\theta_0$ as described in Section A.2.8. These initial values are applied identically in both the actual and counterfactual simulations.

From 1955 onward, the newborn cohort is constructed separately by skill type. High-skilled births from high-skilled and low-skilled parents are computed according to Equation (1.15) in the main paper, low-skilled births follow Equation (1.16).

After calculating the newborn groups, all existing cohorts are advanced by five years using Equation (1.22). High-skilled survival rates are adjusted using the shock parameter $\psi_{i,t}$, as described in Section A.2.6. In contrast, low-skilled survival rates are computed as residual to ensure consistency with the aggregate survival rate. Migration is applied uniformly across skill types but differs between the actual and counterfactual simulations, following the approach outlined in Section A.2.7.

At each step, the share of high-skilled individuals in each cohort $\theta_{i,t}$ is updated recursively using Equation (A.1.46) in Section A.1.7. This process continues in five-year intervals until the year 2015. It produces complete time paths for the population by skill type under both actual and counterfactual scenarios.

## A.3 Figures and tables

Figure A.1: Net aggregate migrants.

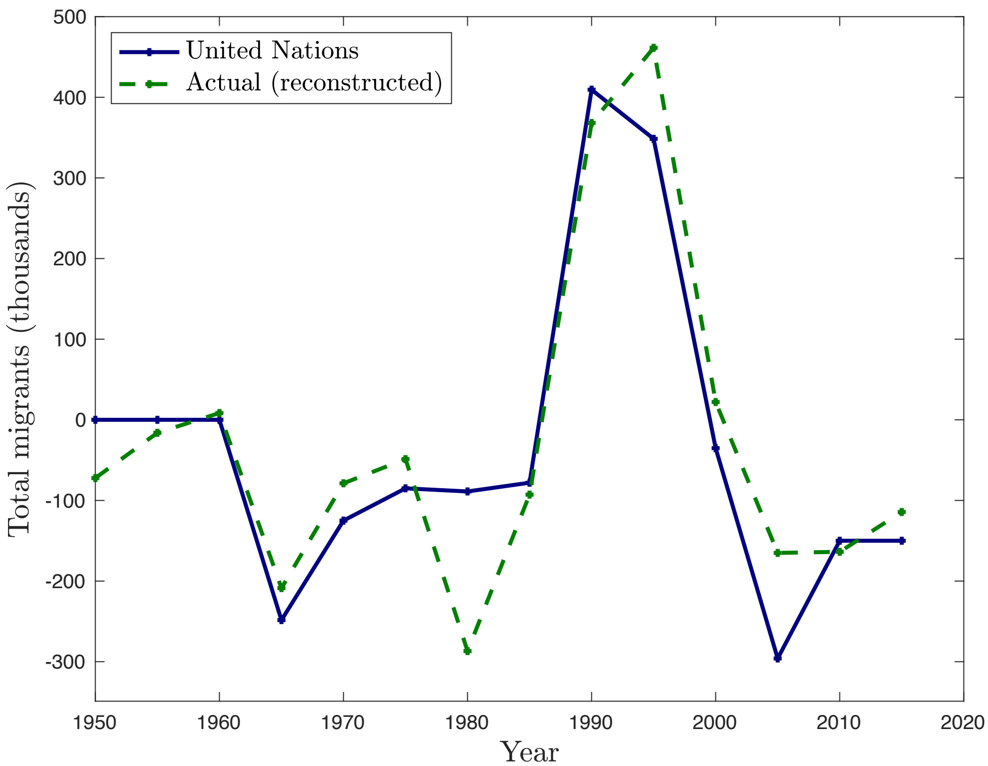
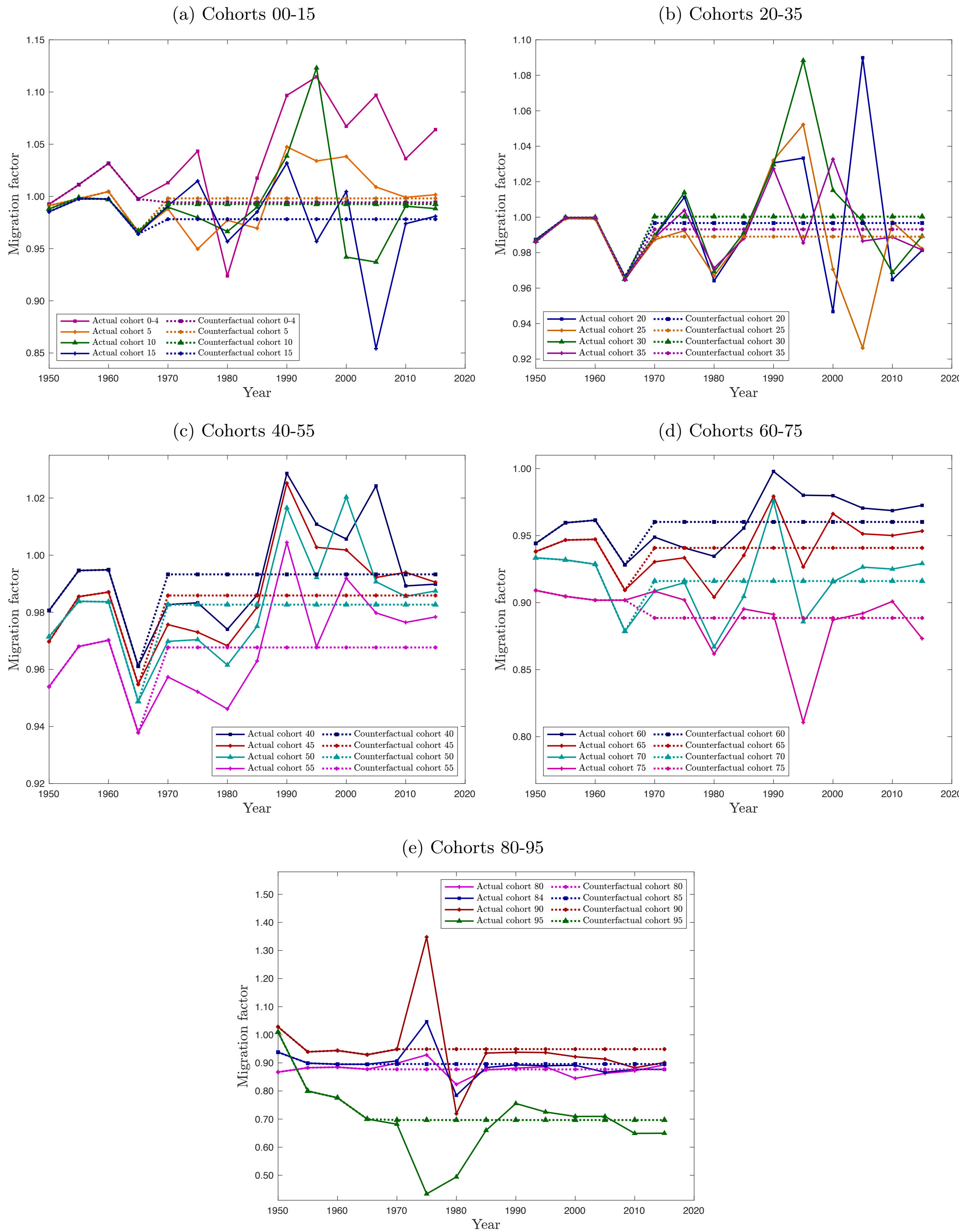


Figure A.2: Actual (reconstructed) vs counterfactual age-specific migration factor.




(a) Cohorts 00-15

(b) Cohorts 20-35

(c) Cohorts 40-55

(d) Cohorts 60-75

(e) Cohorts 80-95

Figure A.3: Actual high- vs low-skilled survival rates.

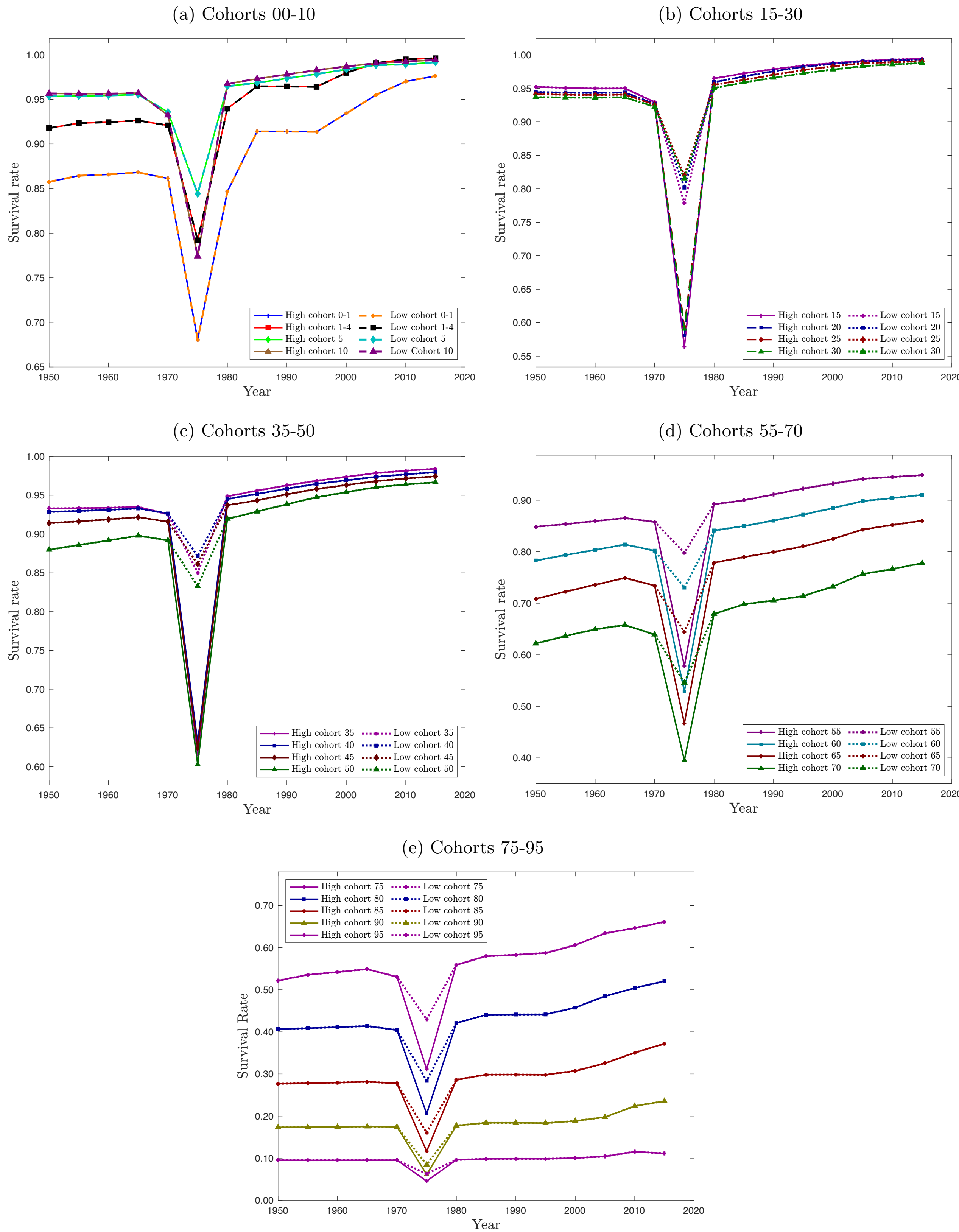


Figure A.4: $\theta_{i,t}$ time path.

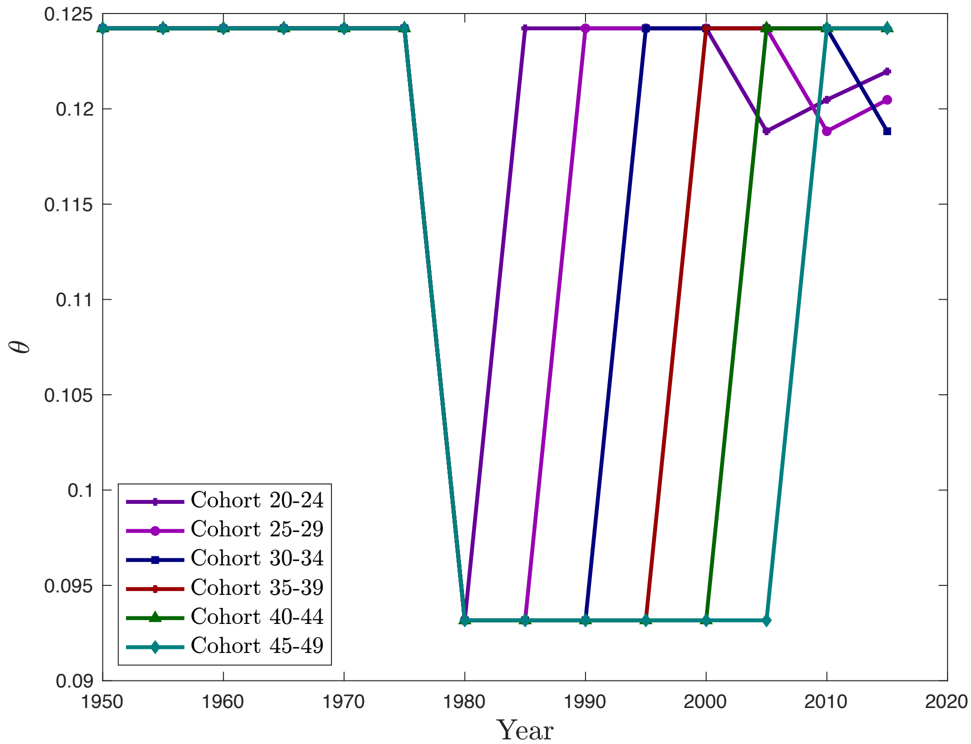



Figure A.5: $\theta_{i,t}$ values by cohort.

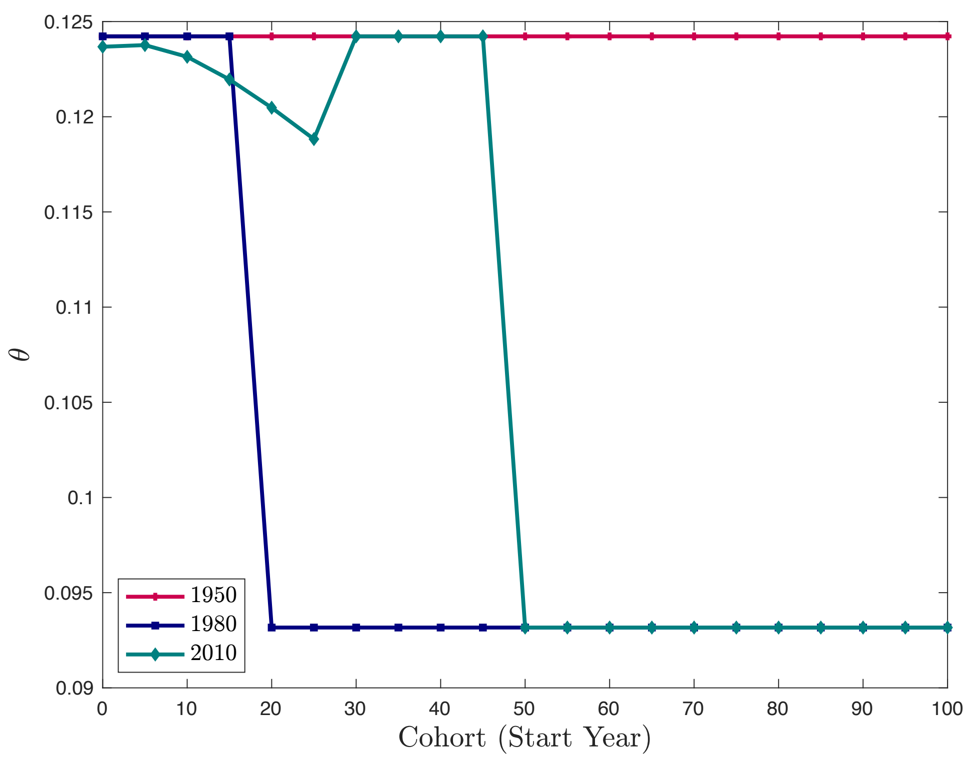


Figure A.6: $\Phi_{i,t}$ time path.

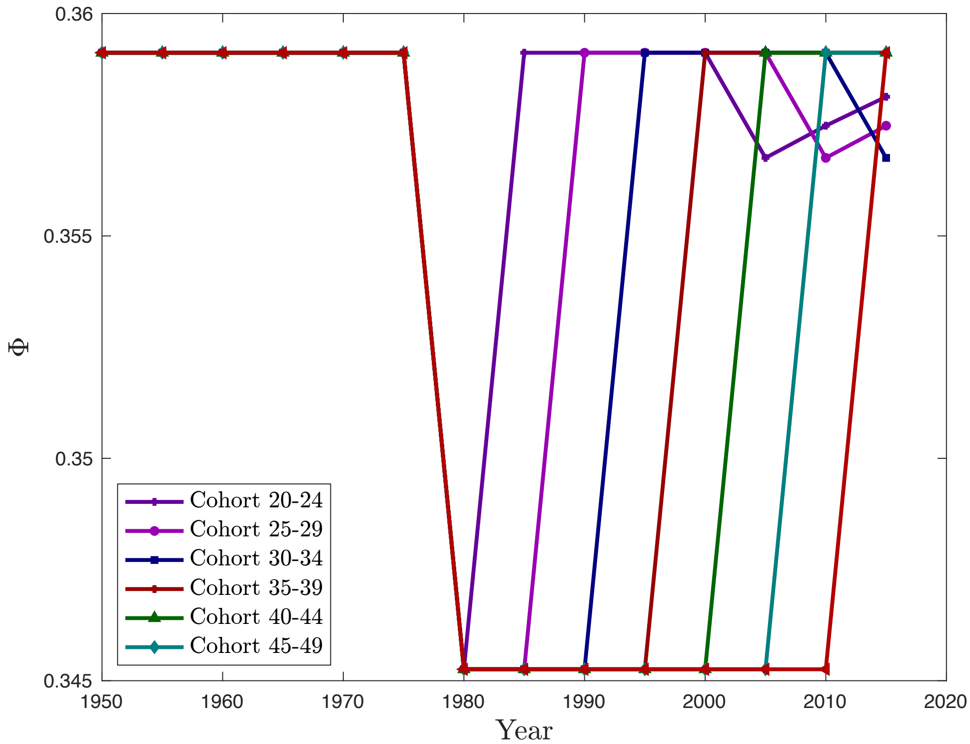



Figure A.7: Gross growth rate of GDP per capita.

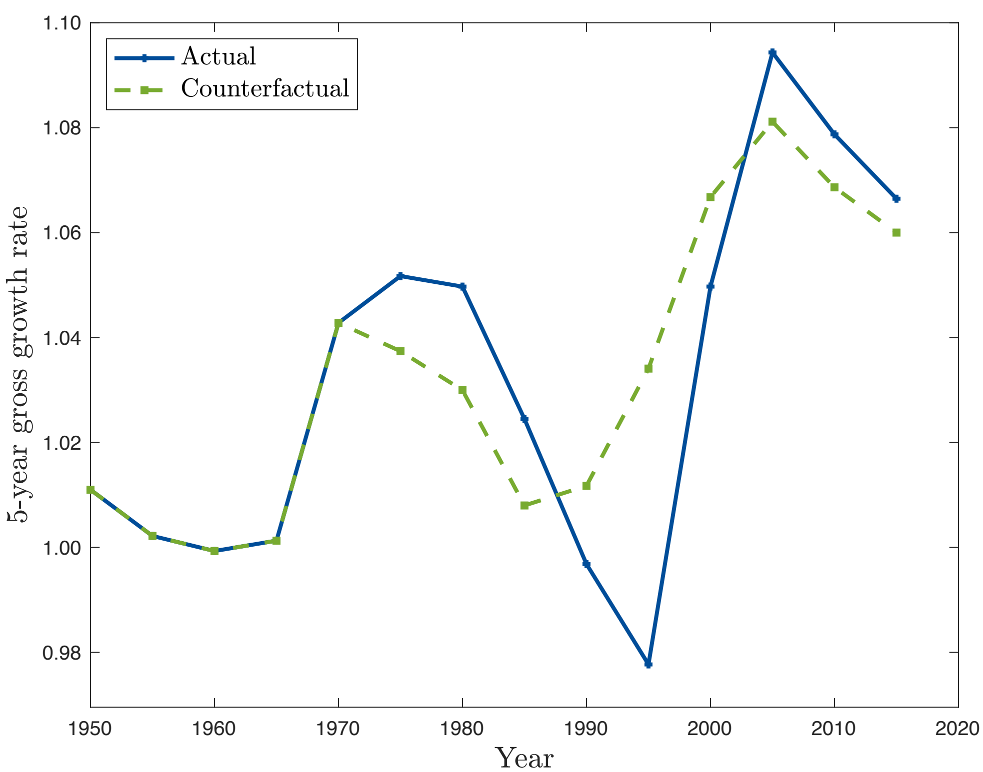


Figure A.8: Gross growth rate of capital.

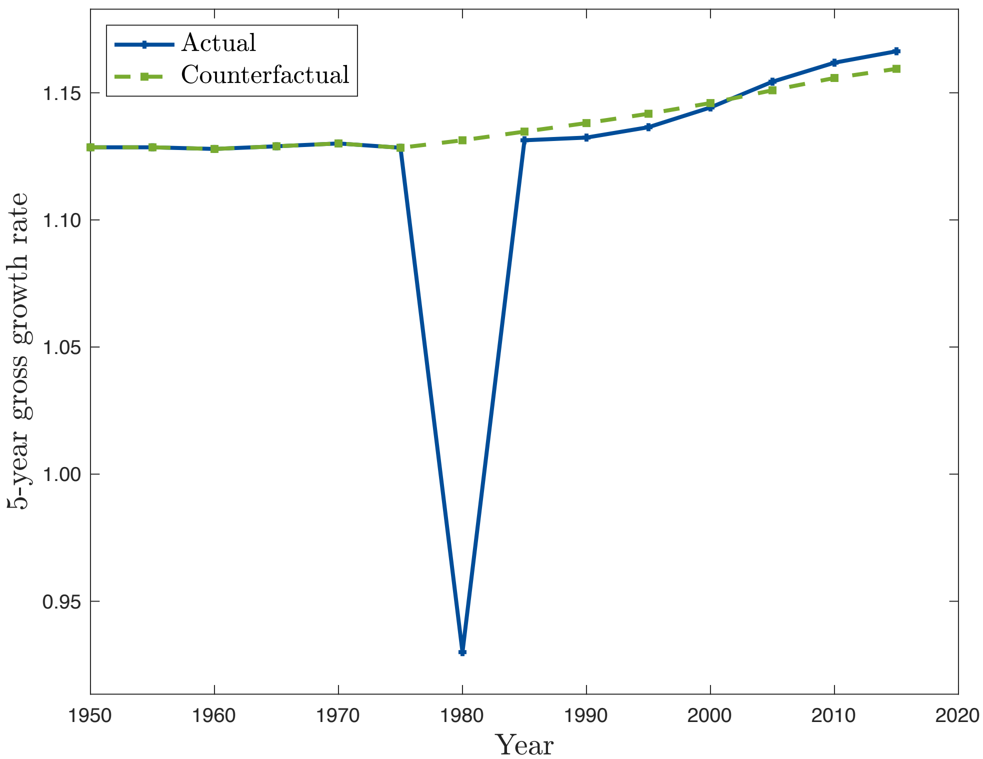



Figure A.9: Gross growth rate of labor.

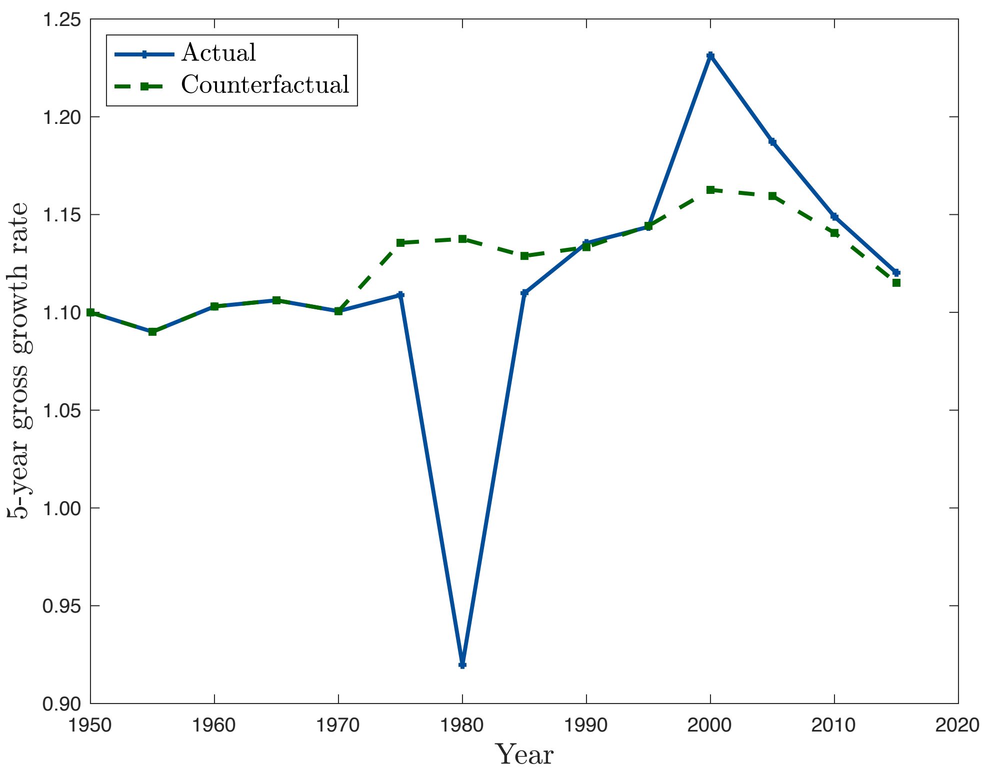


Figure A.10: Gross growth rate of land productivity.

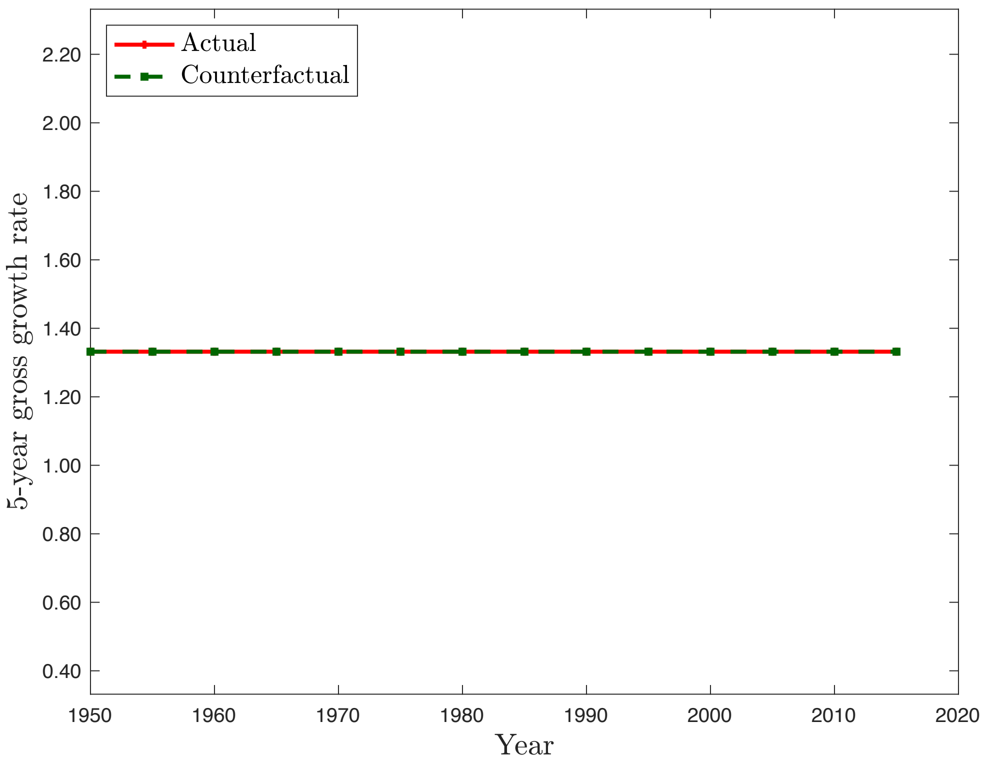



Figure A.11: Relative wages of high- and low-skilled workers (normalized to cohort 15 low-skilled workers).

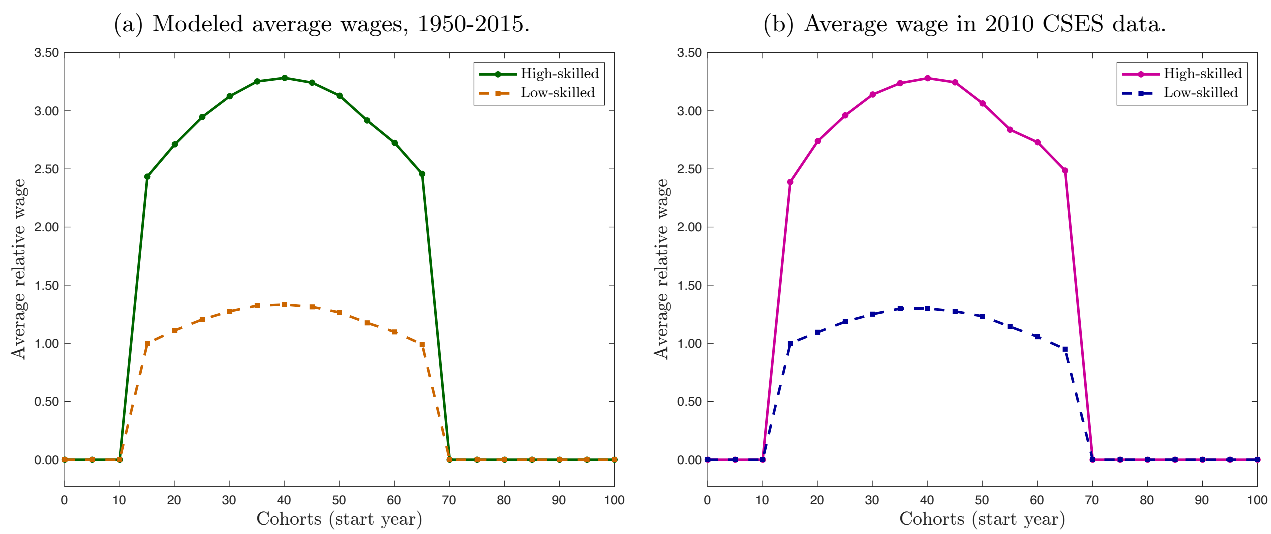


Figure A.12: Relative wages by cohort (normalized to cohort 15).

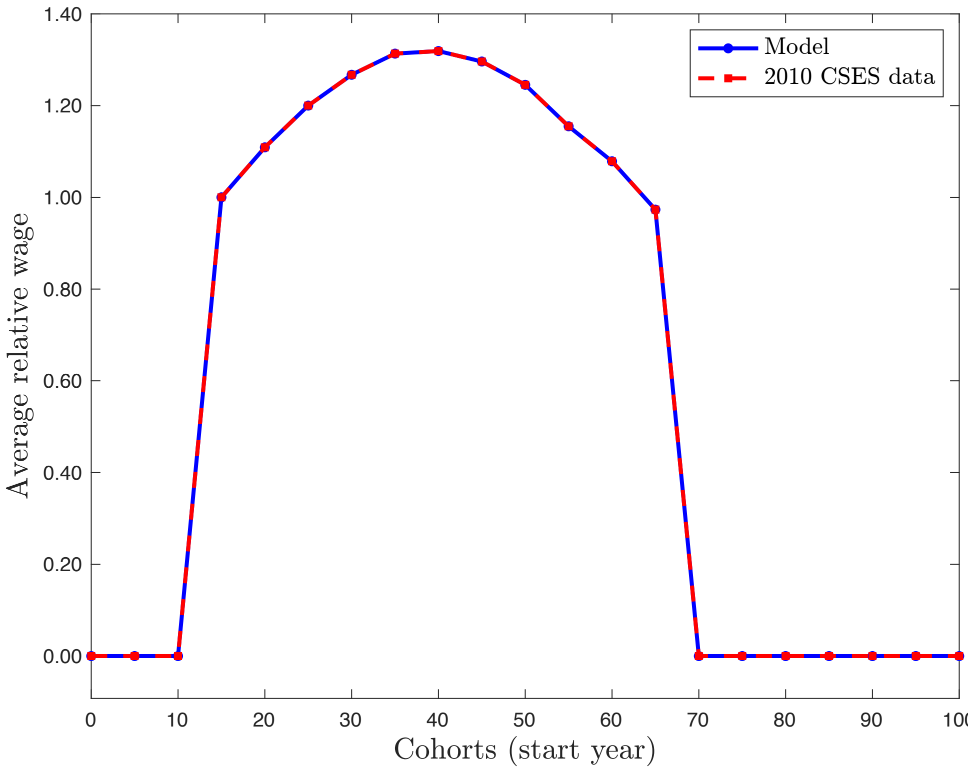



Figure A.13: Average saving rates by cohort.

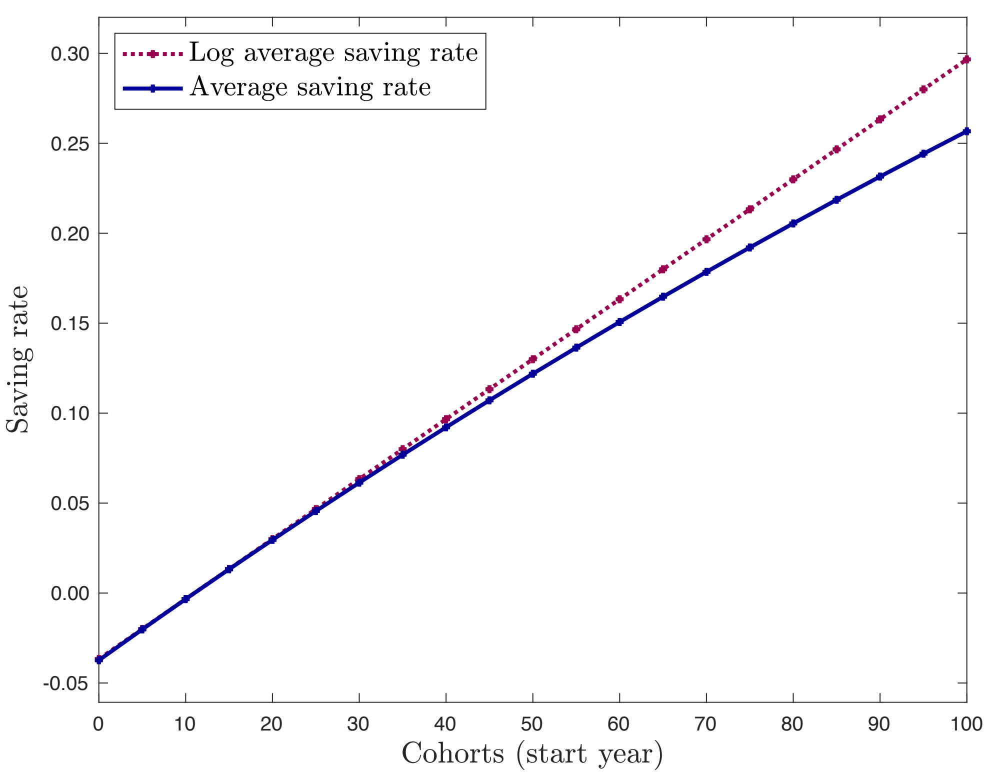


Figure A.14: Cohort-specific saving rates by skill type.

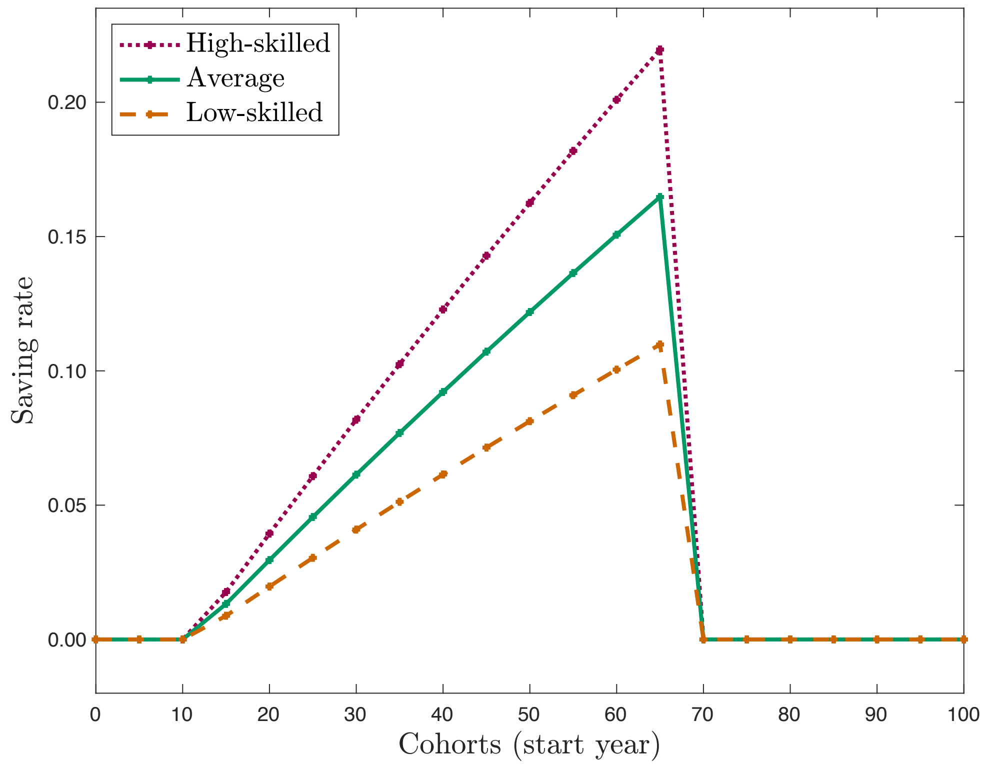
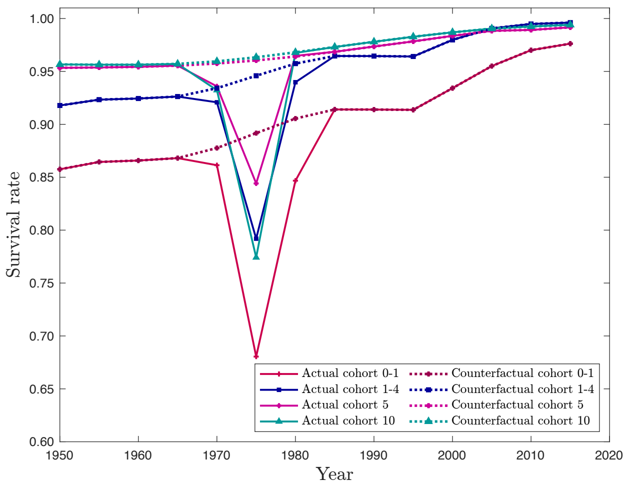
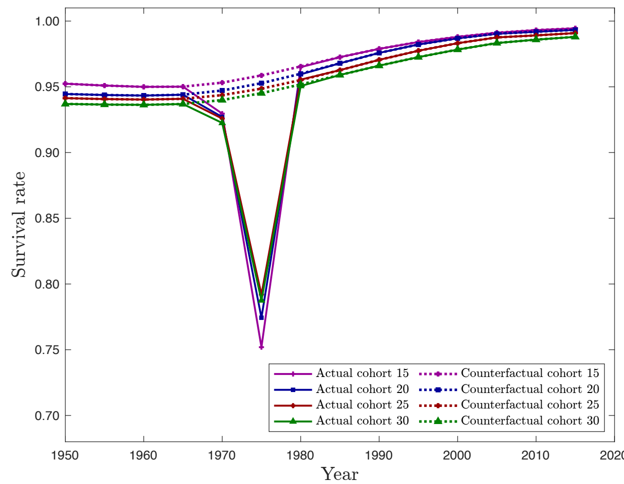


Figure A.15: Actual vs counterfactual age-specific survival rates.

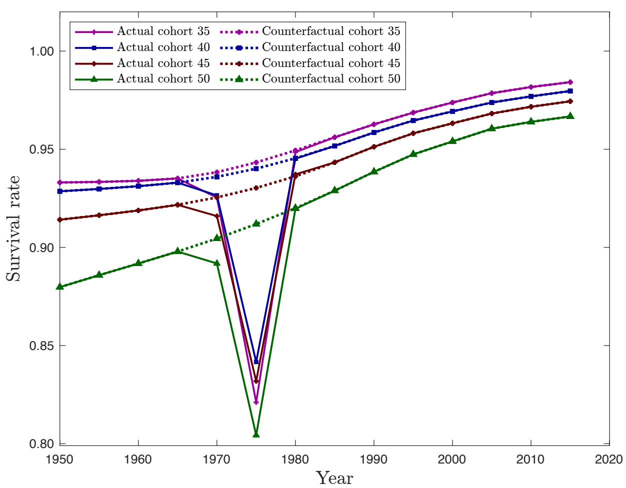
(a) Cohorts 00-10



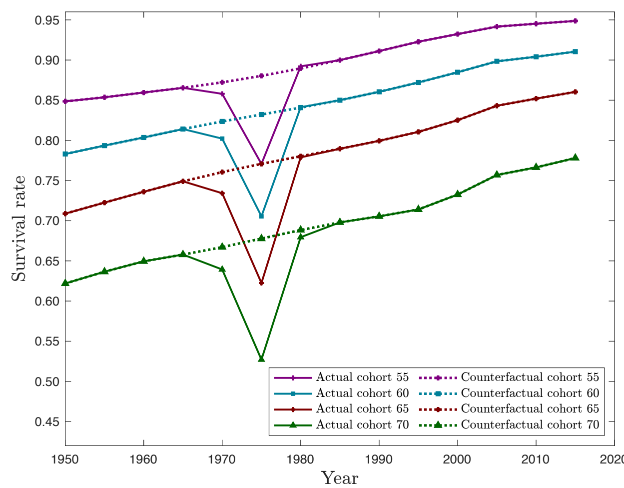

(b) Cohorts 15-30




(c) Cohorts 35-50




(d) Cohorts 55-70




(e) Cohorts 75-90

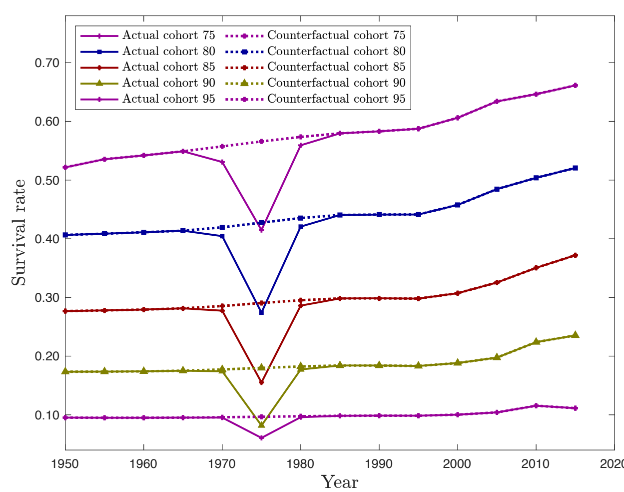


Figure A.16: Counterfactual high- vs low-skilled survival rates.

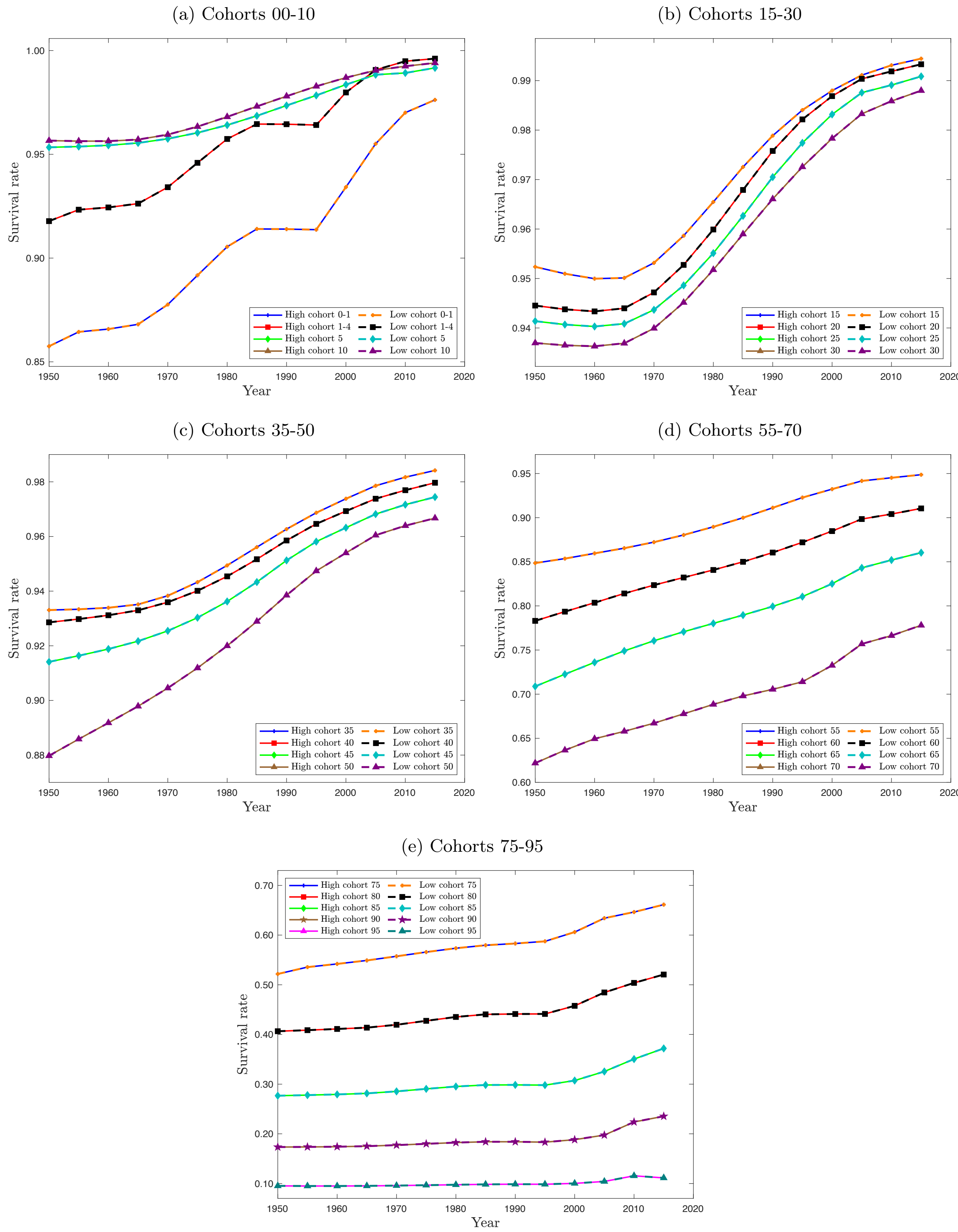


Figure A.17: Age distribution of three selective years.

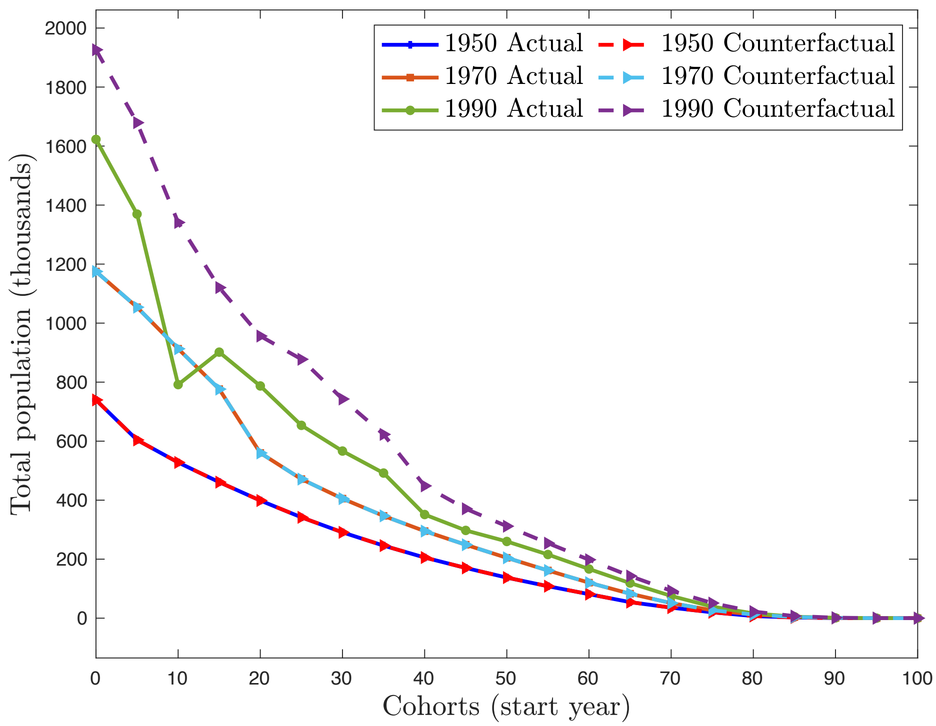



Figure A.18: $\theta_{0,t}$ time path under alternative $\pi$ settings.

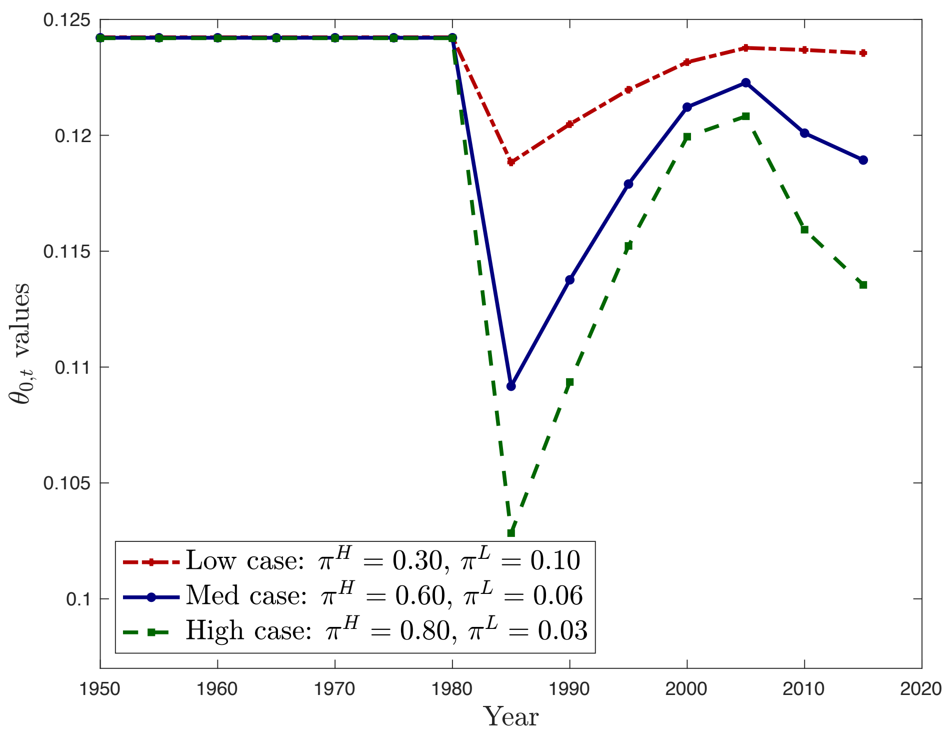


Figure A.19: Total high-skilled population under alternative $\pi$ settings.

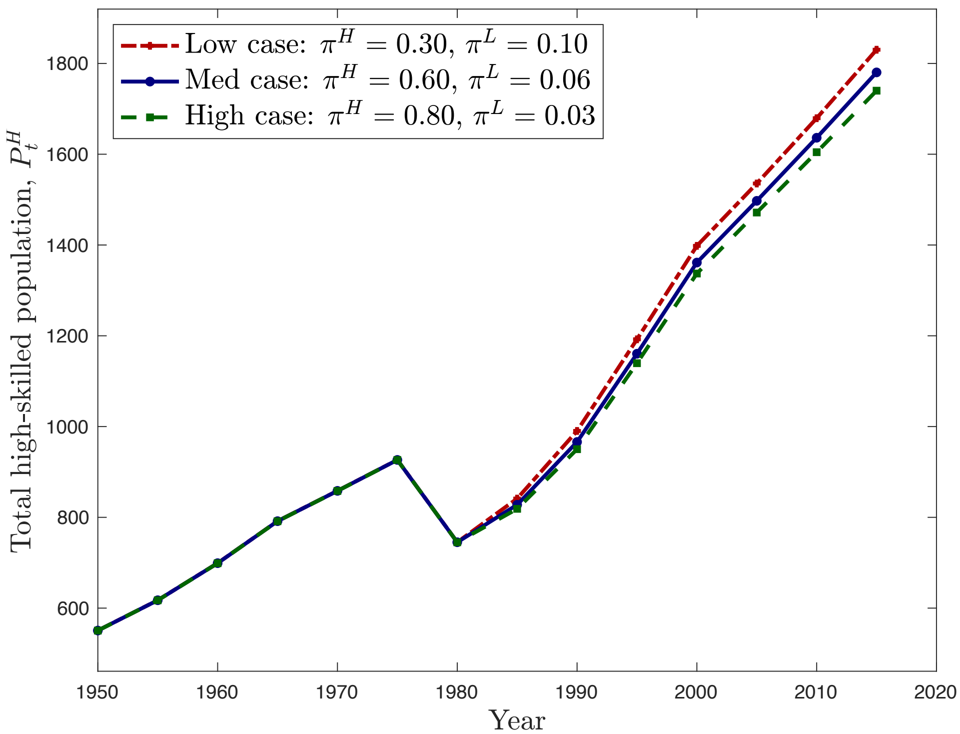



Figure A.20: Total low-skilled population under alternative $\pi$ settings.

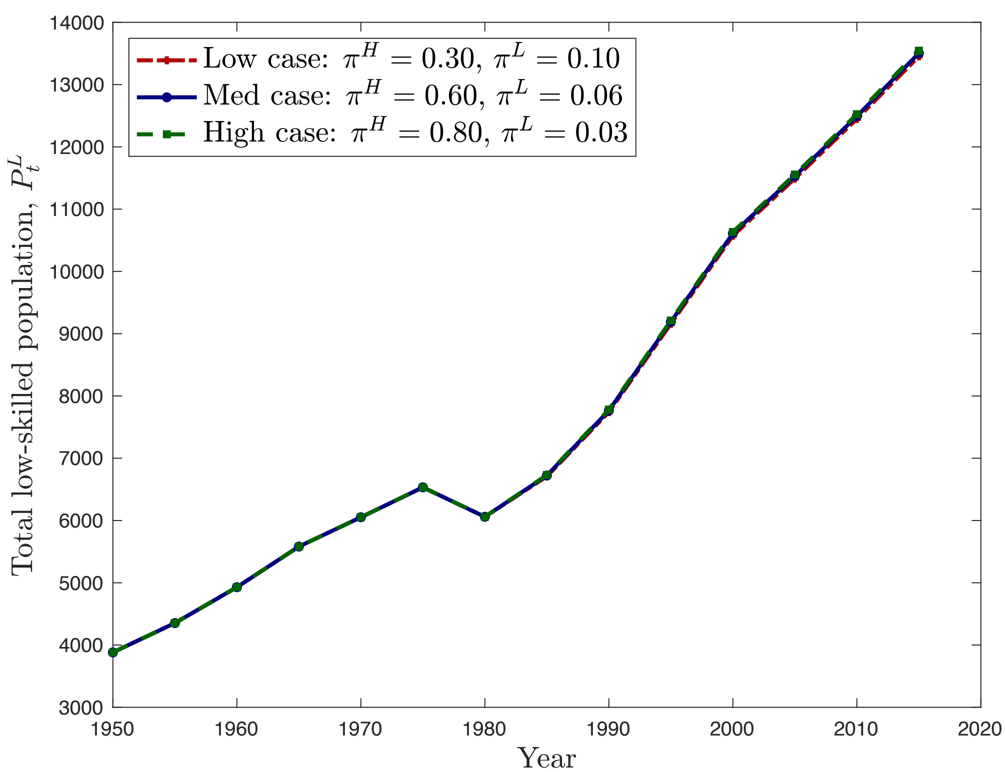


Figure A.21: Relative capital per worker effect: Actual vs. counterfactual.

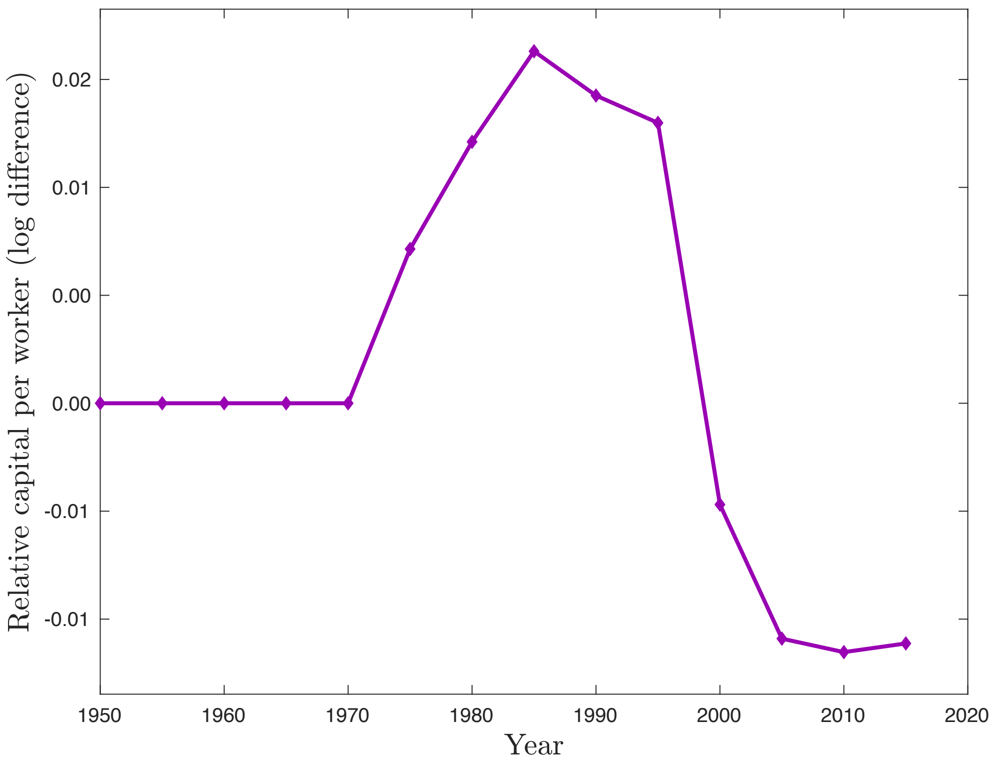



Figure A.22: Relative land per worker effect: Actual vs. counterfactual.

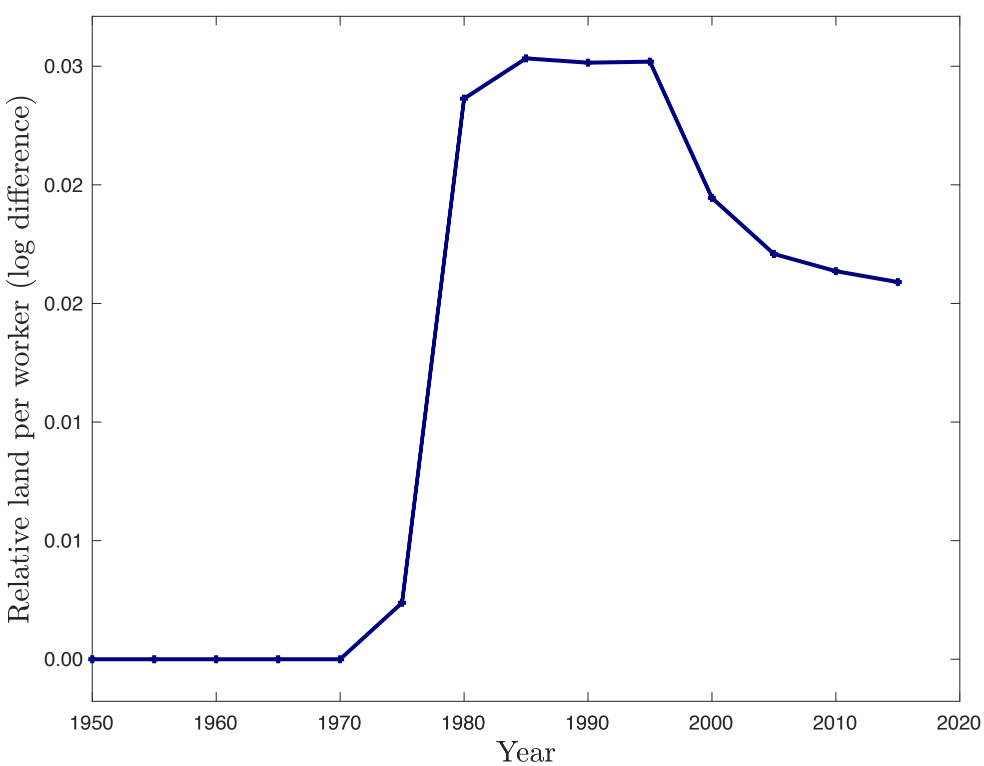


Figure A.23: Relative labor-to-population effect: Actual vs. counterfactual.

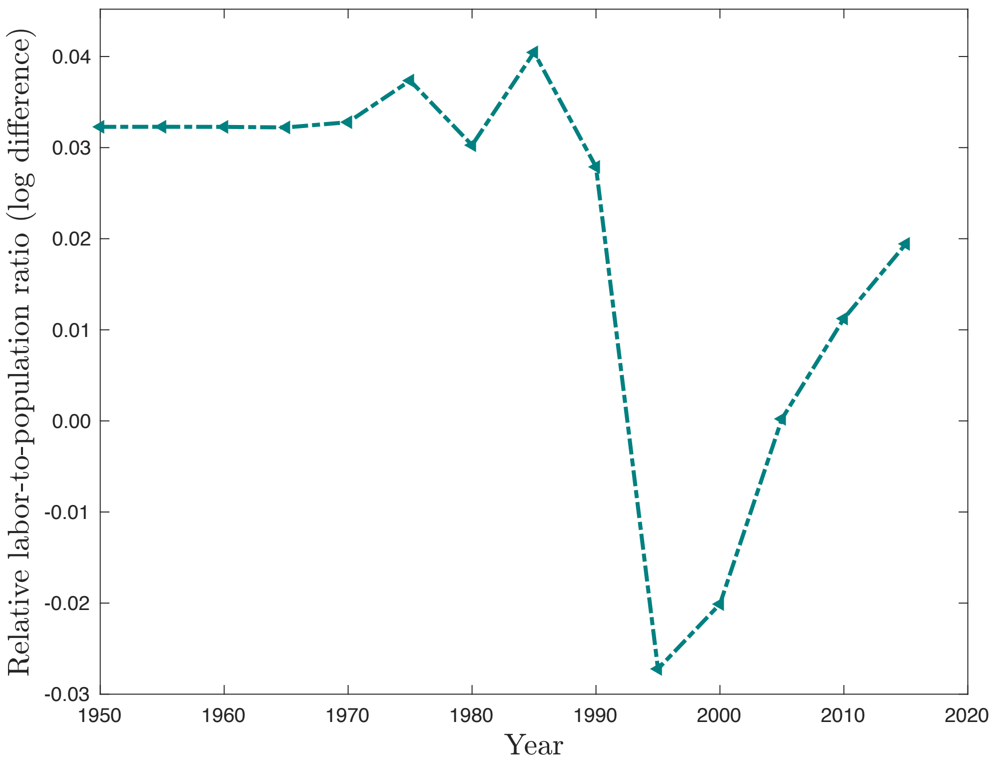



Figure A.24: GDP per capita.

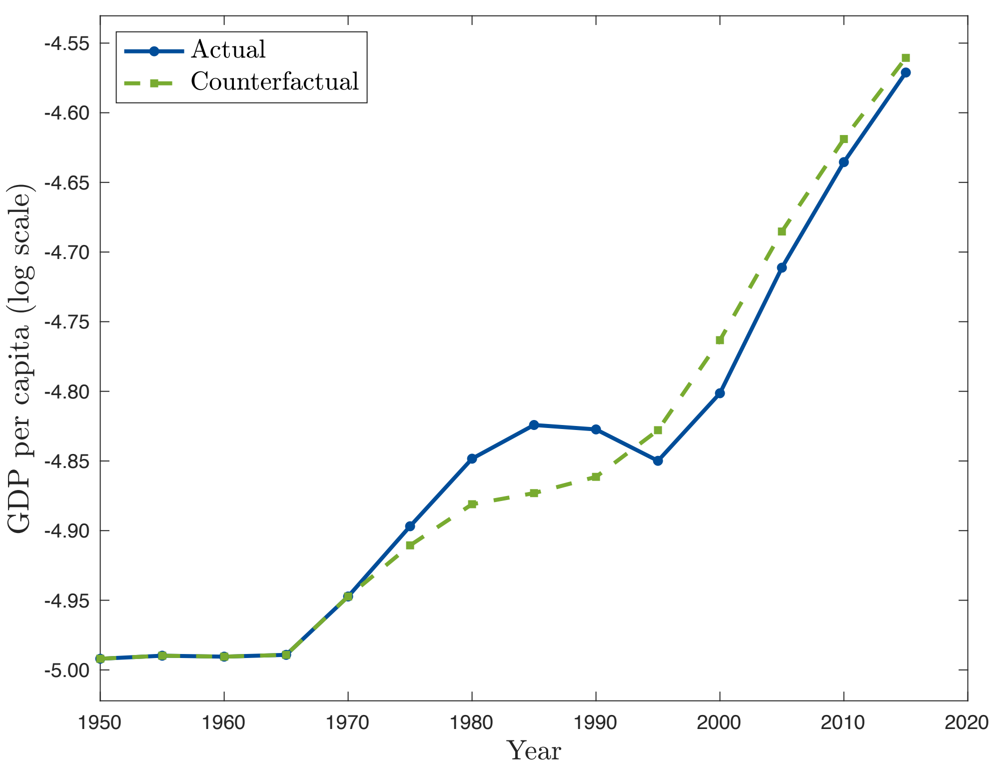


Figure A.25: Capital.

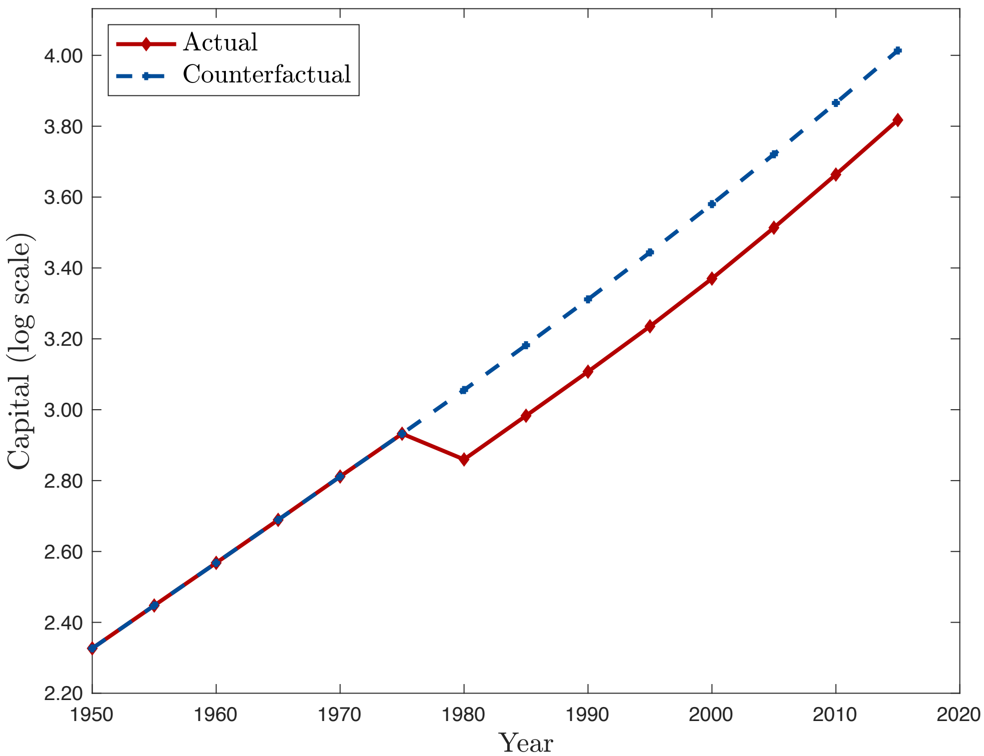



Figure A.26: Relative GDP per capita dropping skill heterogeneity: Special case I.

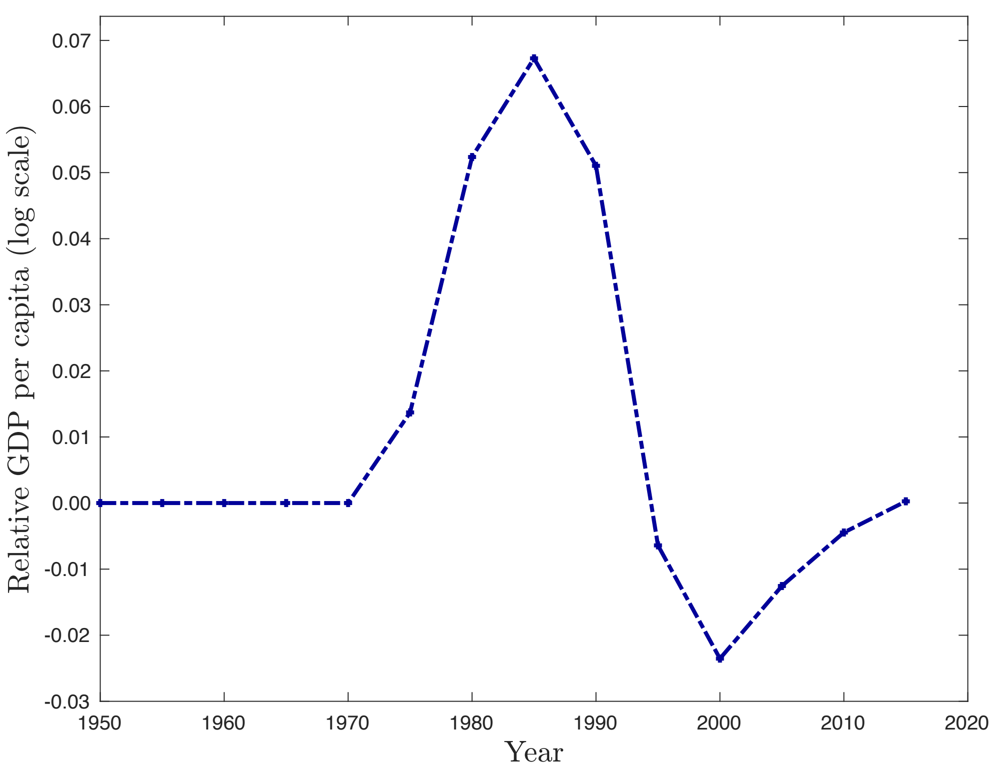


Figure A.27: Relative GDP per capita dropping capital and skill heterogeneity: Special case II.

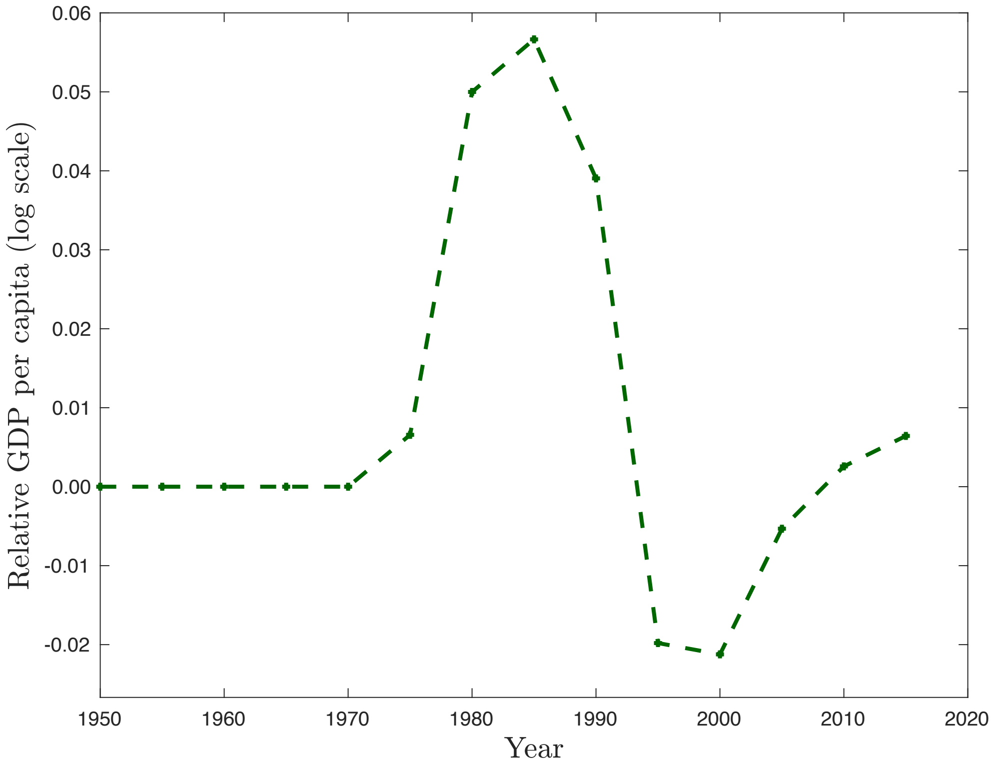


Table A.1: Benchmark parameter values.

| Parameter | Value | Description |
|---|---|---|
| $\theta_0$ | 0.1242 | The initial fraction of potential high-skilled groups |
| $\pi^H$ | 0.295 | The probability of children born as the high-skilled type from high-skilled parents |
| $\pi^L$ | 0.10 | The probability of children born as the high-skilled type from low-skilled parents |
| $\alpha$ | 0.30 | Capital share of output; same as Ashraf et al. (2013) |
| $\beta$ | 0.60 | Labor share of output |
| $\rho$ | 0.90 | The substitutability between the labor supply of different cohorts |
| $\eta$ | 0.90 | The elasticity of substitution between the high-skilled and low-skilled labor |
| $\lambda$ | 0.67 | The relative high-skilled productivity |
| $g_y$ | 0.011 | The initial net growth rate of living standard at 1950 |
| $\gamma_i$ | Table A.3 | The relative productivity of different cohorts $i$ |
| $\sigma_i^H$ | Table A.3 | Saving rates from high-skilled incomes for cohort $i$ |
| $\sigma_i^L$ | Table A.3 | Saving rates from low-skilled incomes for cohort $i$ |
| $\sigma^K$ | 0.0855 | Saving rates from capital incomes; borrowed from Ashraf et al. (2013); Heston et al. (2009) |
| $\sigma^X$ | 0.0855 | Saving rates from land incomes; borrowed from Ashraf et al. (2013); Heston et al. (2009) |
| $\delta$ | 0.07 | Capital depreciation rates; borrowed from Schmitt-Grohé, S., Uribe, M. (2005) |
| $\psi_{i,t}$ | Eq. (A.1.42) | A time-dependent, exogenous survival shock parameter |

Table A.2: Fraction of high-skilled workers by age cohort in 2010 (%).

| Age Group | CSES Data | Reconstructed |
|---|---|---|
| 30 - 34 | 16.62 | 12.42 |
| 35 - 39 | 18.48 | 12.42 |
| 40 - 44 | 14.23 | 12.42 |
| 45 - 49 | 6.73 | 12.42 |
| 50 - 54 | 9.09 | 9.32 |
| 55 - 59 | 6.25 | 9.32 |
| 60 - 64 | 8.89 | 9.32 |

Notes: People whose years of schooling are greater 12 years are categorized as the high-skilled group, and otherwise the low-skilled group in the 2010 CSES. The fraction of high-skilled is defined as $P^H_{i,t}/P_{i,t}$.

Table A.3: Calibrated values of $\gamma_i$ and $\sigma_i^j$ by age cohort.

| Age group | $\gamma_i$ | $\sigma_i^H$ | $\sigma_i^L$ |
|---|---|---|---|
| 15 - 19 | 0.083 | 0.018 | 0.009 |
| 20 - 24 | 0.091 | 0.039 | 0.020 |
| 25 - 29 | 0.097 | 0.061 | 0.030 |
| 30 - 34 | 0.101 | 0.082 | 0.041 |
| 35 - 39 | 0.103 | 0.103 | 0.051 |
| 40 - 44 | 0.102 | 0.123 | 0.061 |
| 45 - 49 | 0.099 | 0.143 | 0.071 |
| 50 - 54 | 0.094 | 0.163 | 0.081 |
| 55 - 59 | 0.085 | 0.182 | 0.091 |
| 60 - 64 | 0.077 | 0.201 | 0.100 |
| 65 - 69 | 0.067 | 0.220 | 0.110 |

*Notes:* $\gamma_i$ denotes the calibrated productivity weights by age cohort. These weights sum to 1 across working-age cohorts (ages 15-69), and are set to 0 outside this range. $\sigma_i^j$ denotes the saving rate from income source $j$'s for cohort $i$, where $j \in \{H, L\}$ represents high-skilled individuals, respectively. These saving rates are assumed to be constant over time.

Table A.4: Total population by age cohort in 2010.

| | Total Population | | | | | |
|---|---|---|---|---|---|---|
| | Data | | Actual | | Counterfactual | |
| Age Group | in 1000s | in % | in 1000s | in % | in 1000s | in % |
| 00 - 04 | 1,658.88 | 11.68 | 1,584.56 | 11.22 | 1,867.20 | 11.46 |
| 05 - 09 | 1,504.21 | 10.59 | 1,504.21 | 10.65 | 1,708.12 | 10.49 |
| 10 - 14 | 1,611.78 | 11.35 | 1,611.78 | 11.41 | 1,767.32 | 10.85 |
| 15 - 19 | 1,658.75 | 11.68 | 1,658.75 | 11.74 | 1,798.92 | 11.04 |
| 20 - 24 | 1,366.69 | 9.62 | 1,366.69 | 9.67 | 1,713.32 | 10.52 |
| 25 - 29 | 1,651.44 | 11.13 | 1,651.44 | 11.69 | 1,518.42 | 9.32 |
| 30 - 34 | 647.10 | 4.56 | 647.10 | 4.58 | 1,203.99 | 7.39 |
| 35 - 39 | 865.23 | 6.09 | 865.23 | 6.12 | 1,004.55 | 6.17 |
| 40 - 44 | 780.47 | 5.50 | 780.47 | 5.52 | 855.01 | 5.25 |
| 45 - 49 | 695.12 | 4.49 | 695.12 | 4.92 | 766.52 | 4.71 |
| 50 - 54 | 503.57 | 3.55 | 503.57 | 3.56 | 634.63 | 3.90 |
| 55 - 59 | 435.46 | 3.07 | 435.46 | 3.08 | 511.36 | 3.14 |
| 60 - 64 | 298.81 | 2.10 | 298.81 | 2.11 | 344.46 | 2.11 |
| 65 - 69 | 219.87 | 1.55 | 219.87 | 1.56 | 251.94 | 1.55 |
| 70 - 74 | 154.05 | 1.08 | 154.05 | 1.09 | 172.97 | 1.06 |
| 75 - 79 | 94.36 | 0.66 | 94.36 | 0.67 | 101.01 | 0.62 |
| 80 - 84 | 40.77 | 0.29 | 40.77 | 0.29 | 47.20 | 0.29 |
| 85 - 89 | 13.24 | 0.09 | 13.24 | 0.09 | 16.09 | 0.10 |
| 90 - 94 | 2.70 | 0.02 | 2.69 | 0.02 | 3.67 | 0.02 |
| 95 - 99 | 0.39 | 0.00 | 0.39 | 0.00 | 0.53 | 0.00 |
| 100 + | 0.02 | 0.00 | 0.02 | 0.00 | 0.03 | 0.00 |
| Total | 14,202.91 | 100 | 14,129.59 | 100 | 16,287.27 | 100 |

Table A.5: Fraction of female population.

| | Reproductive Cohorts | | | | | | |
|---|---|---|---|---|---|---|---|
| Year | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-45 | 45-49 |
| 1950 | 0.495 | 0.495 | 0.495 | 0.496 | 0.497 | 0.501 | 0.506 |
| 1955 | 0.497 | 0.493 | 0.492 | 0.493 | 0.494 | 0.498 | 0.505 |
| 1960 | 0.501 | 0.496 | 0.491 | 0.490 | 0.491 | 0.495 | 0.501 |
| 1965 | 0.503 | 0.501 | 0.495 | 0.490 | 0.490 | 0.493 | 0.498 |
| 1970 | 0.504 | 0.505 | 0.500 | 0.494 | 0.490 | 0.491 | 0.495 |
| 1975 | 0.507 | 0.508 | 0.506 | 0.502 | 0.496 | 0.491 | 0.493 |
| 1980 | 0.549 | 0.550 | 0.542 | 0.540 | 0.525 | 0.500 | 0.493 |
| 1985 | 0.524 | 0.556 | 0.553 | 0.545 | 0.544 | 0.527 | 0.500 |
| 1990 | 0.502 | 0.524 | 0.555 | 0.553 | 0.545 | 0.544 | 0.528 |
| 1995 | 0.510 | 0.504 | 0.526 | 0.556 | 0.554 | 0.547 | 0.546 |
| 2000 | 0.500 | 0.516 | 0.509 | 0.520 | 0.543 | 0.574 | 0.562 |
| 2005 | 0.492 | 0.506 | 0.520 | 0.516 | 0.524 | 0.547 | 0.589 |
| 2010 | 0.485 | 0.505 | 0.509 | 0.521 | 0.517 | 0.526 | 0.540 |
| 2015 | 0.489 | 0.488 | 0.511 | 0.513 | 0.527 | 0.521 | 0.530 |

Table A.6: Total high- and low-skilled population in 2010 CSES data.

| | Low Skill | | High Skill | |
|---|---|---|---|---|
| Age Group | in Number | in Percentage | in Number | in Percentage |
| 15 - 19 | 453 | 98.69 | 6 | 1.31 |
| 20 - 24 | 568 | 90.30 | 61 | 9.70 |
| 25 - 29 | 493 | 76.91 | 148 | 23.09 |
| 30 - 34 | 311 | 83.38 | 62 | 16.62 |
| 35 - 39 | 278 | 81.52 | 63 | 18.48 |
| 40 - 44 | 241 | 85.77 | 40 | 14.23 |
| 45 - 49 | 194 | 93.27 | 14 | 6.73 |
| 50 - 54 | 150 | 90.91 | 15 | 9.09 |
| 55 - 59 | 90 | 93.75 | 6 | 6.25 |
| 60 - 64 | 41 | 91.11 | 4 | 8.89 |
| 65 - 69 | 10 | 76.92 | 3 | 23.08 |
| Total | 2829 | 87.02 | 422 | 12.98 |

Table A.7: Relative GDP and GDP per capita (2010): Baseline model.

| High-skill child probability | Relative GDP | Relative GDP per capita |
|---|---|---|
| $\pi^H = 0.30$ and $\pi^L = 0.10$ | 0.8532 | 0.9836 |
| $\pi^H = 0.60$ and $\pi^L = 0.06$ | 0.8510 | 0.9810 |
| $\pi^H = 0.80$ and $\pi^L = 0.03$ | 0.8495 | 0.9793 |

*Notes:* Relative GDP measures the total output in 2010 of actual GDP relative to counterfactual scenarios, with relative GDP per capita defined analogously. The production function uses factor shares of $\alpha = 0.3$ for capital, $\beta = 0.6$ for labor, and the remaining share $1-\alpha-\beta = 0.1$ for land. During the Pol Pot regime (1975–1979) in the actual scenario, saving rates from all three income sources (capital, labor, and land) are set to zero. The substitution parameters in the nested effective labor input, $\rho$ and $\eta$, are both set to 0.9. The values of $\pi^H$ and $\pi^L$ in each case are chosen to ensure that $\theta_0$ is approximately equal to 12.42 percent.

# Appendix B

## Chapter 2 Appendix

### Figures

Figure B.1: Total number of births in urban areas by birth year.

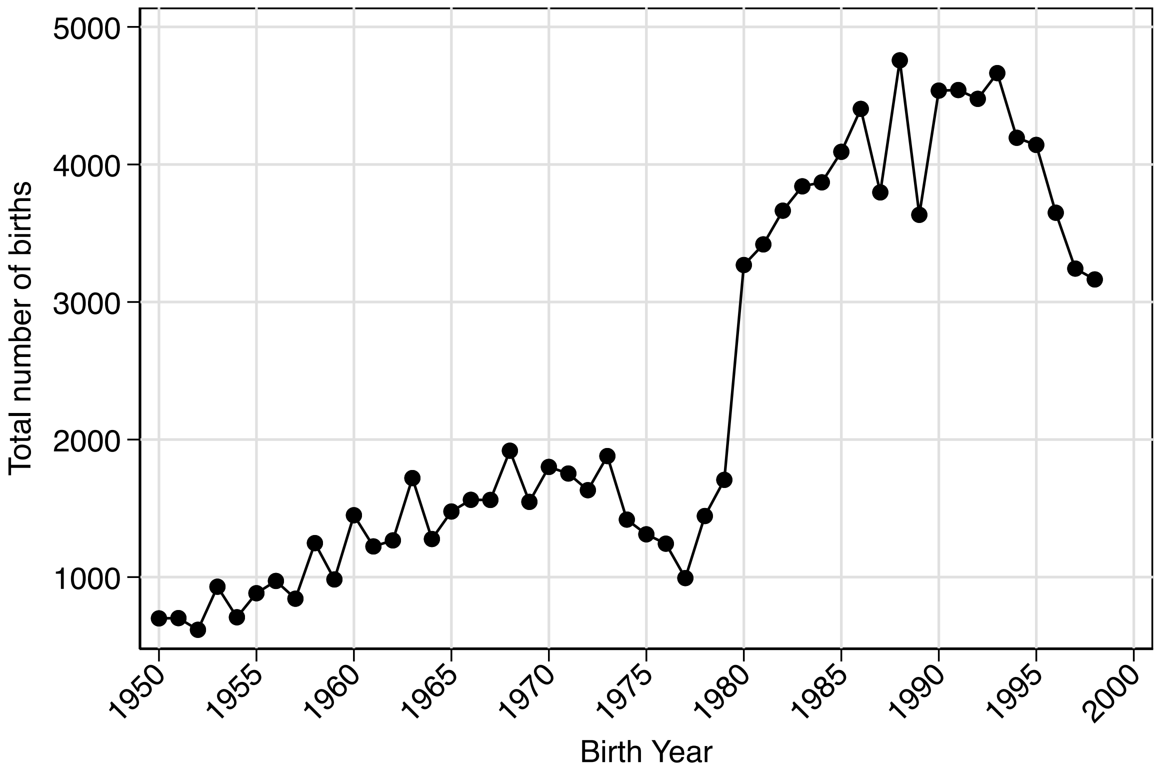
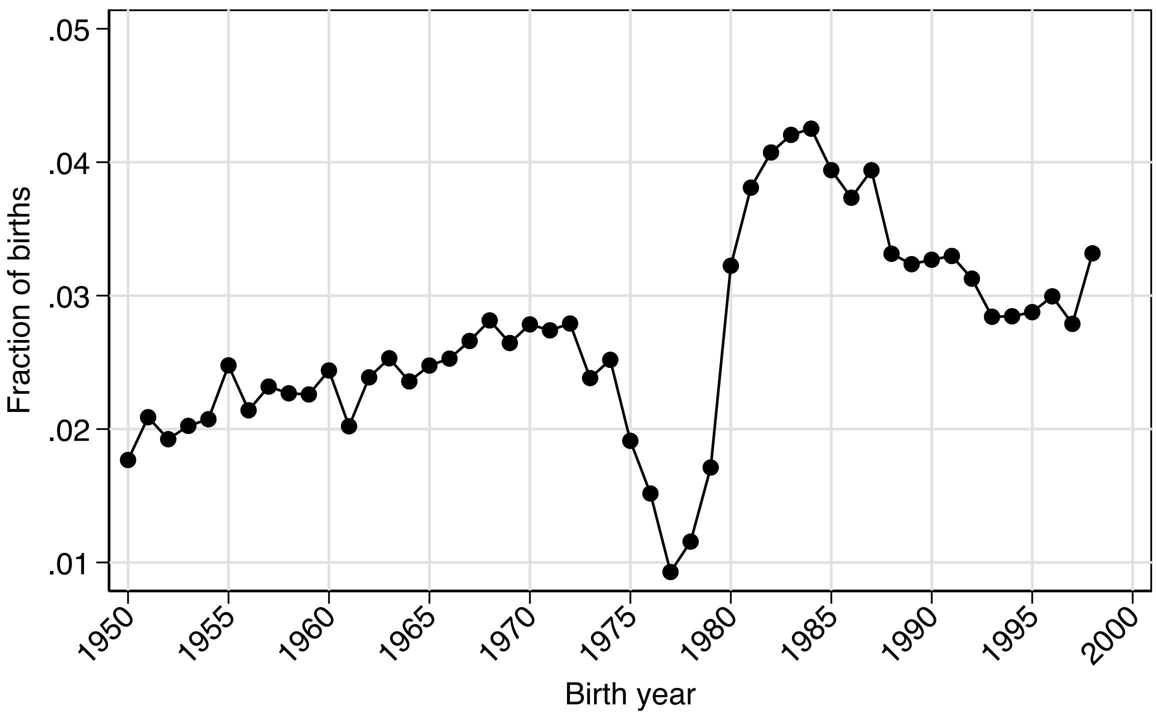


Figure B.2: Fraction of births in Phnom Penh by birth year.




Note: There are 7 districts in Phnom Penh city, and 4 of them are classified as urban districts, based on General Population Census of Cambodia 1998, Final Census Results.

Figure B.3: Predicted years of schooling based on district-specific cubic time trends.

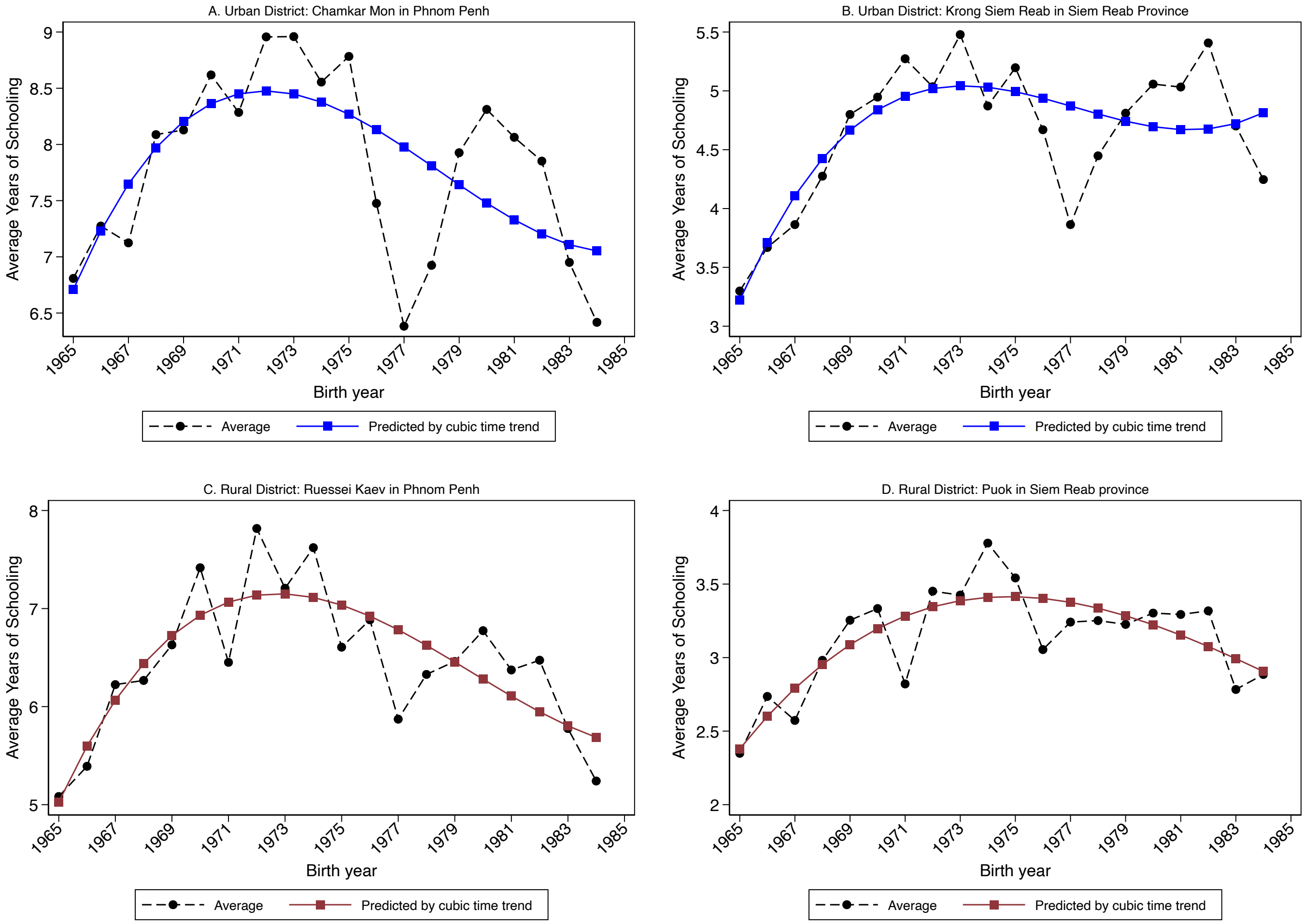


Figure B.4: Heatmap of cross-district correlations between urban proxy and urban dummy.

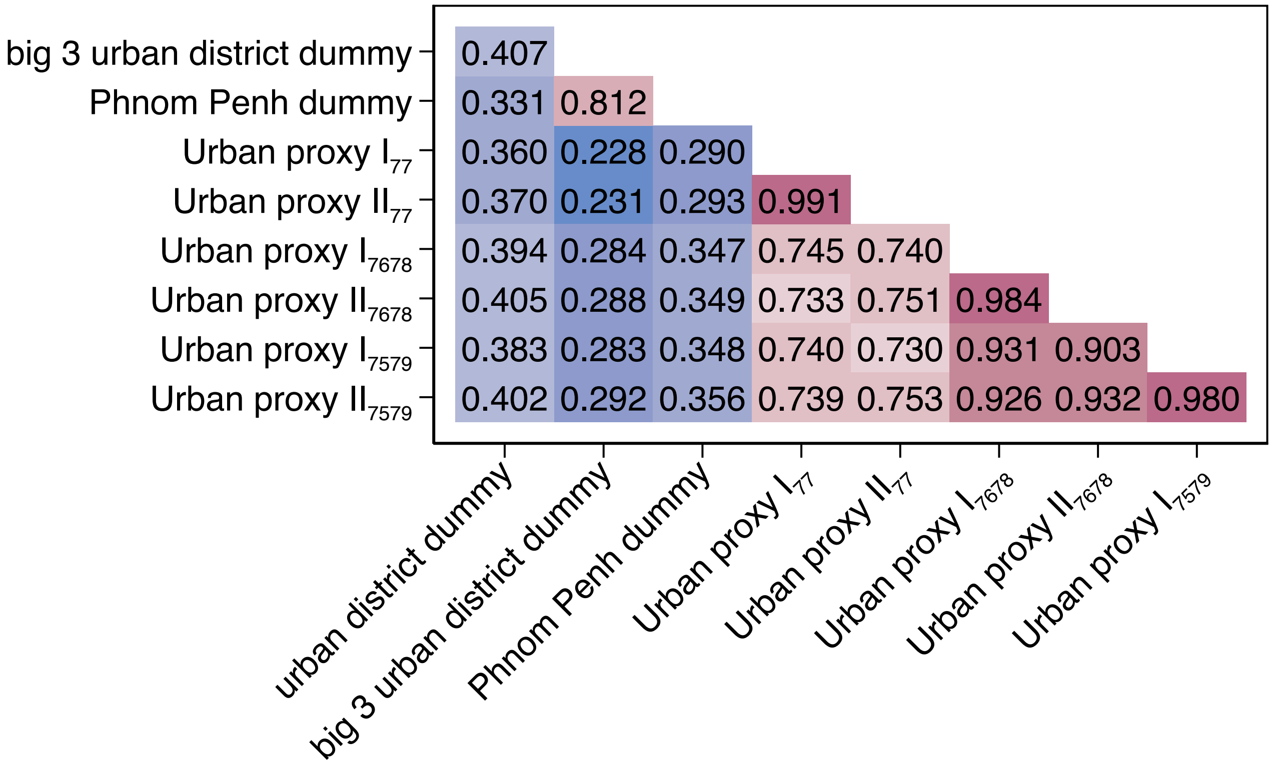
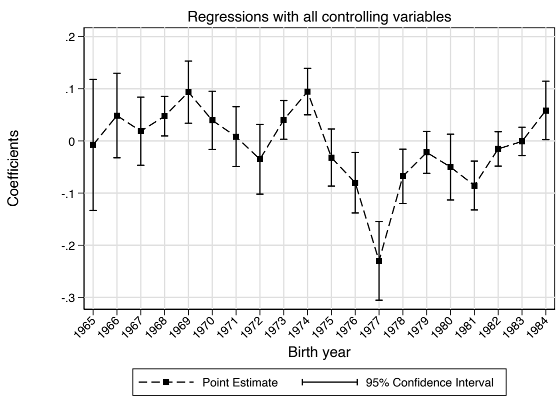
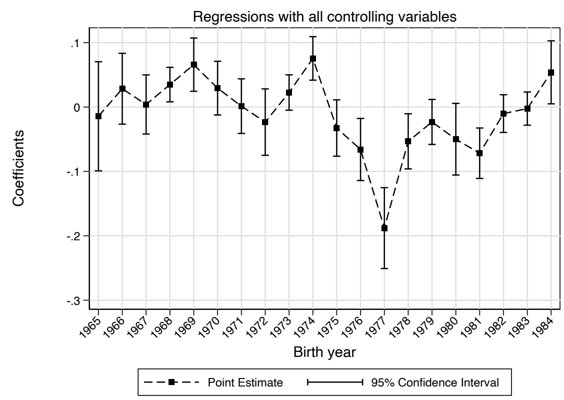


Figure B.5: Estimated years of schooling and cohort dummy.

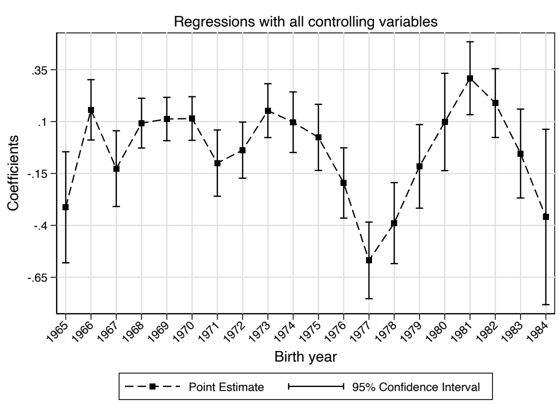
(a) Interaction cohort and urban proxy I.



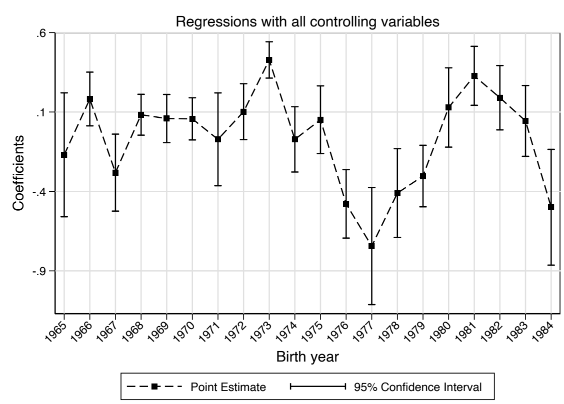

(b) Interaction cohort and urban proxy II.




(c) Interaction cohort and urban dummy.




(d) Interaction cohort and big 3 urban dummy.




(e) Interaction cohort and Phnom Penh dummy.

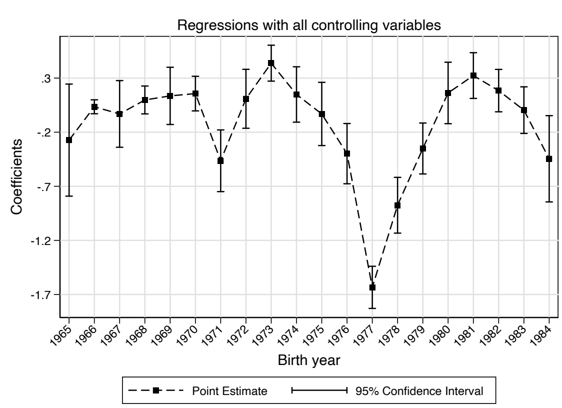


Figure B.6: Average height of females by birth year.

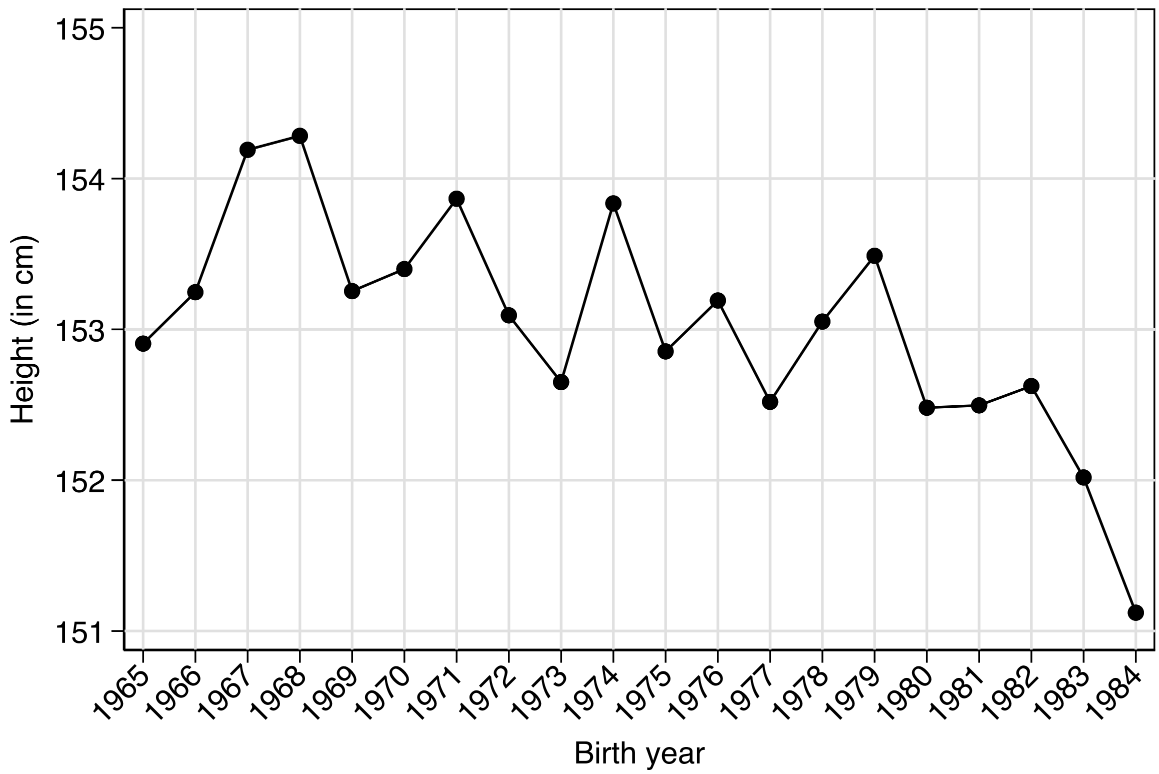



Figure B.7: Average weight of females by birth year.

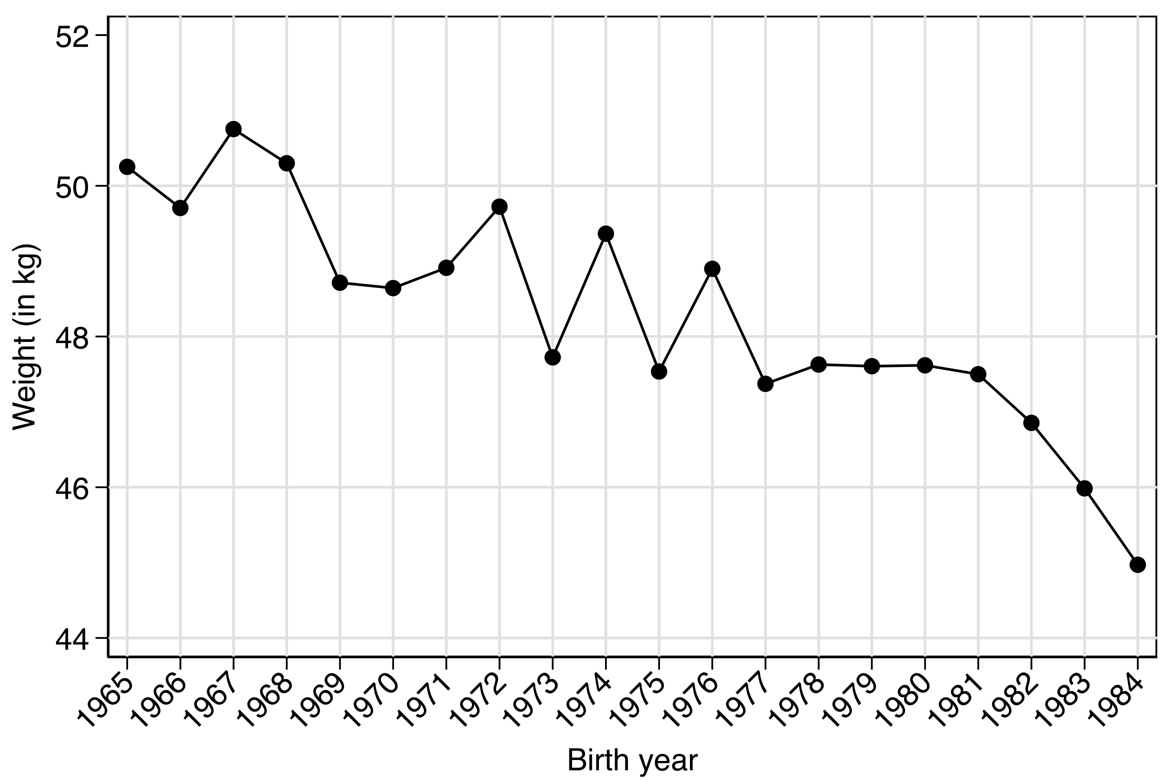


Figure B.8: Average years of schooling for females by birth year.

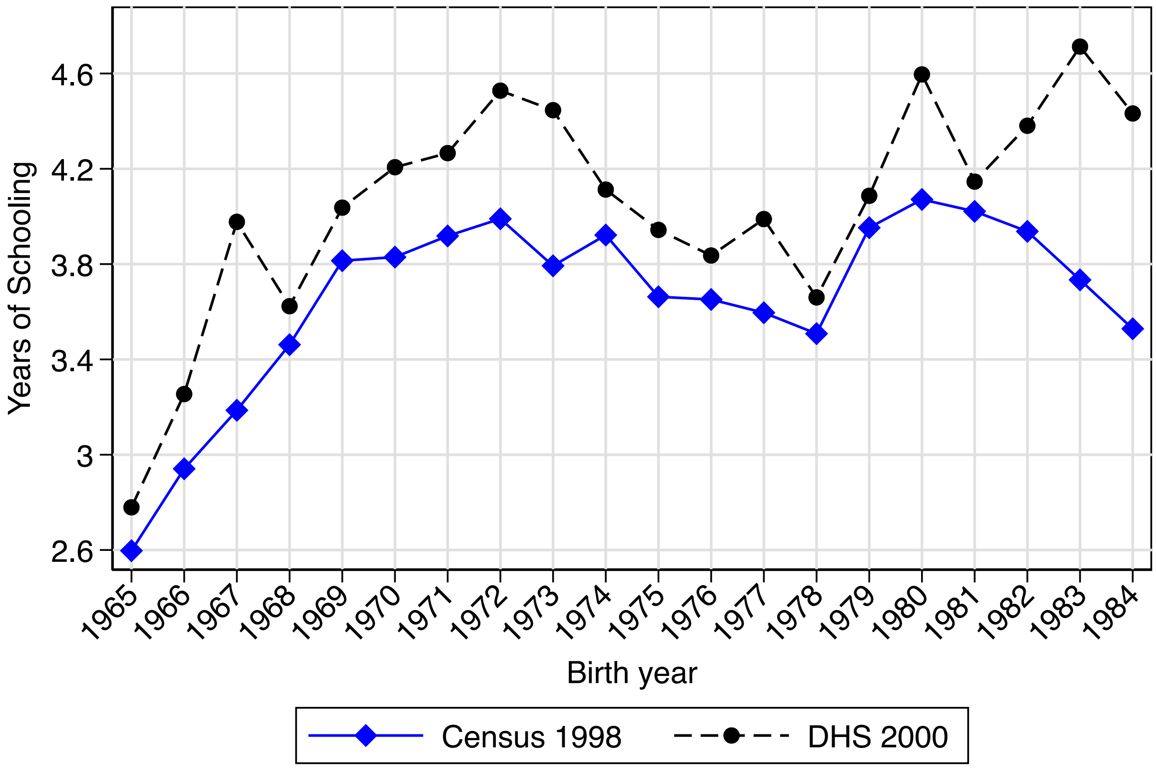
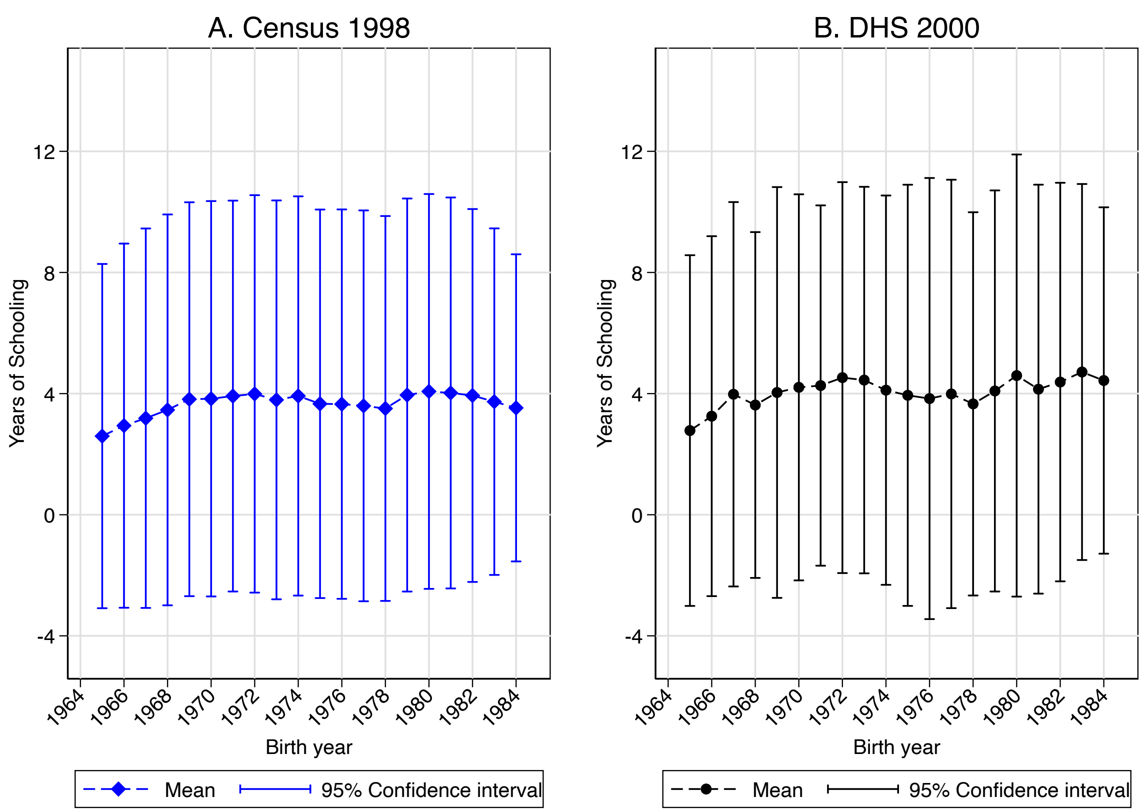



Figure B.9: Average years of schooling for females by birth year.



# Tables

Table B.1: Summary statistics of variables at the individual level.

| | Mean | Std. Error | Min. | Max. | Median | Obs. |
|---|---|---|---|---|---|---|
| Years of schooling | 4.33 | 3.41 | 0 | 13 | 4.00 | 384,044 |
| Cohort dummy$_{77}$ | 0.03 | 0.18 | 0 | 1 | 0.00 | 384,332 |
| Cohort dummy$_{7678}$ | 0.11 | 0.32 | 0 | 1 | 0.00 | 384,332 |
| Cohort dummy$_{7579}$ | 0.20 | 0.40 | 0 | 1 | 0.00 | 384,332 |
| Urban dummy | 0.11 | 0.31 | 0 | 1 | 0.00 | 384,332 |
| Big 3 urban dummy | 0.05 | 0.22 | 0 | 1 | 0.00 | 384,332 |
| Phnom Penh dummy | 0.03 | 0.17 | 0 | 1 | 0.00 | 384,332 |
| Urban proxy $I_{77}$ | 96.69 | 0.84 | 91 | 100 | 96.59 | 384,332 |
| Urban proxy $I_{7678}$ | 88.75 | 2.25 | 79 | 100 | 88.55 | 384,332 |
| Urban proxy $I_{7579}$ | 80.18 | 3.27 | 65 | 92 | 79.78 | 384,332 |
| Urban proxy $II_{77}$ | 95.83 | 1.04 | 88 | 100 | 95.72 | 384,332 |
| Urban proxy $II_{7678}$ | 85.82 | 2.74 | 72 | 100 | 85.40 | 384,332 |
| Urban proxy $II_{7579}$ | 75.03 | 3.93 | 53 | 91 | 74.41 | 384,332 |
| Female | 0.52 | 0.50 | 0 | 1 | 1.00 | 384,332 |
| Single | 0.54 | 0.50 | 0 | 1 | 1.00 | 384,332 |
| Married | 0.44 | 0.50 | 0 | 1 | 0.00 | 384,332 |
| Divorced | 0.02 | 0.13 | 0 | 1 | 0.00 | 384,332 |
| Widowed | 0.01 | 0.09 | 0 | 1 | 0.00 | 384,332 |
| Number of children | 0.85 | 1.36 | 0 | 9 | 0.00 | 384,332 |

*Notes:* Data source is extracted from Minnesota Population Center, Integrated Public Use Microdata Series - International (IPUMS-I) 1998. Observations include individuals born between 1965 and 1984. The urban proxy $I_{77}$, for example, is constructed as $(1 - \frac{PopBorn_{77}}{Totpop_{65-84}}) \times 100$. Similarly, the urban proxy $II_{77}$ is defined as $(1 - \frac{PopBorn_{77}}{Totpop_{70-84}}) \times 100$. All urban proxy variables are measured in percentage terms. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban districts within Phnom Penh. The definition of the urban district is based on the General Population Census of Cambodia 1998, Final Census Results. In total, 31 out of 183 districts are classified as urban. For more details on urban districts, see Table B.28.

Table B.2: Cross-district correlations conditional on birth year.

| Variable being correlated | Average years of schooling | Average wealth index score |
|---|---|---|
| Urban proxy $\mathrm{I}_{77}$ | 0.312*** | 0.434*** |
| Urban proxy $\mathrm{II}_{77}$ | 0.316*** | 0.456*** |
| Urban proxy $\mathrm{I}_{76-78}$ | 0.503*** | 0.461*** |
| Urban proxy $\mathrm{II}_{76-78}$ | 0.508*** | 0.482*** |
| Urban proxy $\mathrm{I}_{75-79}$ | 0.412*** | 0.476*** |
| Urban proxy $\mathrm{II}_{75-79}$ | 0.420*** | 0.507*** |
| urban district dummy | 0.381*** | 0.560*** |
| big 3 urban district dummy | 0.374*** | 0.567*** |
| Phnom Penh dummy | 0.378*** | 0.623*** |

*Notes:* The urban proxy $\mathrm{I}_b$ is defined as $(1 - \frac{PopBorn_b}{Totpop_{65-84}}) \times 100$, where $b$ refers to specific birth years: $1975 - 1979$, $1976 - 1978$, and 1977. Here, $Totpop_{65-84}$ denotes the total number of births between 1965 and 1984. The urban proxy $\mathrm{II}_b$ is defined as $(1 - \frac{PopBorn_b}{Totpop_{70-84}}) \times 100$, where $Totpop_{70-84}$ is the total number of births between 1970 and 1984. Average years of schooling and wealth index score are calculated conditional on birth year. All correlations are measured across 181 districts.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.3: Summary statistics of variables at the household level.

| | Mean | Std. Error | Min. | Max. | Median | Obs. |
|---|---|---|---|---|---|---|
| Wealth Index Score | -0.11 | 0.86 | -1 | 4 | -0.49 | 67,605 |
| Cohort dummy$_{77}$ | 0.02 | 0.15 | 0 | 1 | 0.00 | 72,791 |
| Cohort dummy$_{7678}$ | 0.07 | 0.26 | 0 | 1 | 0.00 | 72,791 |
| Cohort dummy$_{7579}$ | 0.13 | 0.34 | 0 | 1 | 0.00 | 72,791 |
| Urban dummy | 0.08 | 0.27 | 0 | 1 | 0.00 | 72,791 |
| Big 3 urban dummy | 0.03 | 0.17 | 0 | 1 | 0.00 | 72,791 |
| Phnom Penh dummy | 0.01 | 0.11 | 0 | 1 | 0.00 | 72,791 |
| Urban proxy $I_{77}$ | 97.81 | 0.85 | 80 | 100 | 97.80 | 72,791 |
| Urban proxy $I_{7678}$ | 92.69 | 1.92 | 67 | 100 | 92.74 | 72,791 |
| Urban proxy $I_{7579}$ | 87.08 | 2.83 | 60 | 100 | 86.90 | 72,791 |
| Urban proxy $II_{77}$ | 95.96 | 1.54 | 50 | 100 | 95.95 | 72,788 |
| Urban proxy $II_{7678}$ | 86.54 | 3.00 | 50 | 100 | 86.72 | 72,788 |
| Urban proxy $II_{7579}$ | 76.19 | 4.08 | 0 | 100 | 75.82 | 72,788 |
| Female | 0.18 | 0.38 | 0 | 1 | 0.00 | 72,791 |
| Single | 0.08 | 0.27 | 0 | 1 | 0.00 | 72,791 |
| Married | 0.88 | 0.33 | 0 | 1 | 1.00 | 72,791 |
| Divorced | 0.03 | 0.16 | 0 | 1 | 0.00 | 72,791 |
| Widowed | 0.02 | 0.13 | 0 | 1 | 0.00 | 72,791 |
| Number of children | 1.90 | 1.44 | 0 | 9 | 2.00 | 72,791 |

*Notes:* The data source is from Minnesota Population Center's Integrated Public Use Microdata Series - International (IPUMS-I) 1998. Observations are households with individuals born between 1965 and 1984. The urban proxy $\mathrm{I}_{77}$, for example, is constructed as $(1 - \frac{PopBorn_{77}}{Totpop_{65-84}}) \times 100$. Similarly, the urban proxy $\mathrm{II}_{77}$ is defined as $(1 - \frac{PopBorn_{77}}{Totpop_{70-84}}) \times 100$. All urban proxy variables are measured in percentage terms. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban district birthplaces located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy is an indicator for four urban districts within Phnom Penh. The definition of an urban district is based on the General Population Census of Cambodia 1998: Final Census Results. In total, 31 out of 183 districts are classified as urban. For more details on urban districts, see Appendix Table B.28.

Table B.4: Years of schooling and cohort dummy$_{7678}$: Interaction of cohort dummy$_{7678}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{7678}$ × urban proxy I$_{7678}$ | -0.05** | -0.05** | -0.05** | -0.05** | -0.05** | -0.08*** |
| | (0.024) | (0.024) | (0.022) | (0.022) | (0.022) | (0.016) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel B | | | | | | |
| Cohort dummy$_{7678}$ × urban proxy II$_{7678}$ | -0.04** | -0.04** | -0.05** | -0.04** | -0.04** | -0.07*** |
| | (0.020) | (0.019) | (0.018) | (0.018) | (0.018) | (0.014) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel C | | | | | | |
| Cohort dummy$_{7678}$ × urban dummy | -0.27** | -0.27** | -0.28*** | -0.24** | -0.25** | -0.47*** |
| | (0.099) | (0.094) | (0.096) | (0.103) | (0.097) | (0.114) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel D | | | | | | |
| Cohort dummy$_{7678}$ × big 3 urban dummy | -0.31** | -0.31** | -0.31** | -0.26* | -0.26** | -0.65*** |
| | (0.122) | (0.112) | (0.125) | (0.129) | (0.120) | (0.163) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel E | | | | | | |
| Cohort dummy$_{7678}$ × Phnom Penh dummy | -0.48 | -0.48 | -0.50* | -0.41 | -0.44 | -0.97*** |
| | (0.289) | (0.282) | (0.270) | (0.268) | (0.260) | (0.254) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 384,044. The urban proxy I$_{7678}$ is constructed as $(1 - \frac{PopBorn_{7678}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7678}$ is defined as $(1 - \frac{PopBorn_{7678}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.5: Years of schooling and cohort dummy$_{7579}$: Interaction of cohort dummy$_{7579}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{7579}$ × urban proxy I$_{7579}$ | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 | -0.04*** |
| | (0.017) | (0.017) | (0.016) | (0.016) | (0.016) | (0.011) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel B | | | | | | |
| Cohort dummy$_{7579}$ × urban proxy II$_{7579}$ | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 | -0.04*** |
| | (0.014) | (0.014) | (0.013) | (0.013) | (0.013) | (0.009) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel C | | | | | | |
| Cohort dummy$_{7579}$ × urban dummy | -0.13 | -0.13 | -0.15 | -0.09 | -0.11 | -0.40*** |
| | (0.098) | (0.098) | (0.098) | (0.106) | (0.106) | (0.111) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel D | | | | | | |
| Cohort dummy$_{7579}$ × big 3 urban dummy | -0.10 | -0.11 | -0.15 | -0.05 | -0.08 | -0.56*** |
| | (0.118) | (0.111) | (0.102) | (0.119) | (0.109) | (0.123) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Panel E | | | | | | |
| Cohort dummy$_{7579}$ × Phnom Penh dummy | -0.15 | -0.15 | -0.20 | -0.08 | -0.11 | -0.71*** |
| | (0.178) | (0.177) | (0.150) | (0.159) | (0.154) | (0.128) |
| $R^2$ | 0.14 | 0.18 | 0.15 | 0.16 | 0.19 | 0.20 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 384,044. The urban proxy I$_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.

$^*$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.6: Wealth index score and cohort dummy$_{7678}$: Interaction of cohort dummy$_{7678}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{7678}$ × urban proxy I$_{7678}$ | -0.01*** | -0.01*** | -0.02*** | -0.01*** | -0.01*** | -0.00 |
| | (0.004) | (0.004) | (0.005) | (0.004) | (0.005) | (0.006) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel B | | | | | | |
| Cohort dummy$_{7678}$ × urban proxy II$_{7678}$ | -0.01** | -0.01*** | -0.01** | -0.01** | -0.01** | 0.00 |
| | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) | (0.004) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel C | | | | | | |
| Cohort dummy$_{7678}$ × urban dummy | -0.11 | -0.11 | -0.15* | -0.08 | -0.12* | -0.20* |
| | (0.072) | (0.073) | (0.074) | (0.069) | (0.071) | (0.096) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel D | | | | | | |
| Cohort dummy$_{7678}$ × big 3 urban dummy | -0.33*** | -0.33*** | -0.37*** | -0.30*** | -0.34*** | -0.43** |
| | (0.096) | (0.098) | (0.106) | (0.097) | (0.104) | (0.149) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel E | | | | | | |
| Cohort dummy$_{7678}$ × Phnom Penh dummy | -0.43 | -0.44 | -0.44 | -0.37 | -0.39 | -0.55 |
| | (0.328) | (0.332) | (0.358) | (0.328) | (0.350) | (0.362) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 67,603. The urban proxy I$_{7678}$ is constructed as $(1-\frac{PopBorn_{7678}}{Totpop_{65-84}})\times 100$. The urban proxy II$_{7678}$ is defined as $(1-\frac{PopBorn_{7678}}{Totpop_{70-84}})\times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.7: Wealth index score and cohort $\text{dummy}_{7579}$: Interaction of cohort $\text{dummy}_{7579}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban proxy $\text{I}_{7579}$ | -0.01 | -0.01 | -0.01 | -0.00 | -0.01 | 0.01 |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.006) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel B | | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban proxy $\text{II}_{7579}$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | 0.01 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel C | | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban dummy | -0.06 | -0.06 | -0.08 | -0.03 | -0.06 | -0.11 |
| | (0.048) | (0.049) | (0.057) | (0.047) | (0.055) | (0.102) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel D | | | | | | |
| Cohort $\text{dummy}_{7579}$ × big 3 urban dummy | -0.12 | -0.13 | -0.16 | -0.09 | -0.13 | -0.10 |
| | (0.150) | (0.150) | (0.154) | (0.156) | (0.158) | (0.230) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Panel E | | | | | | |
| Cohort $\text{dummy}_{7579}$ × Phnom Penh dummy | -0.08 | -0.09 | -0.12 | -0.02 | -0.05 | 0.00 |
| | (0.391) | (0.393) | (0.403) | (0.405) | (0.410) | (0.519) |
| $R^2$ | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 67,603. The urban proxy $\text{I}_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy $\text{II}_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.8: Female years of schooling and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.14*** | -0.14*** | -0.14*** | -0.13*** | -0.19*** |
| | (0.036) | (0.036) | (0.035) | (0.035) | (0.039) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel B | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.11*** | -0.11*** | -0.11*** | -0.11*** | -0.16*** |
| | (0.028) | (0.028) | (0.027) | (0.027) | (0.032) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel C | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.27*** | -0.25*** | -0.26*** | -0.25*** | -0.39*** |
| | (0.065) | (0.064) | (0.067) | (0.067) | (0.093) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel D | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.58*** | -0.56*** | -0.55*** | -0.53*** | -0.87*** |
| | (0.081) | (0.081) | (0.086) | (0.090) | (0.127) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel E | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -1.15*** | -1.15*** | -1.12*** | -1.10*** | -1.62*** |
| | (0.106) | (0.105) | (0.106) | (0.108) | (0.141) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 198,477. The urban proxy I$_{77}$ is constructed as $(1-\frac{PopBorn_{77}}{Totpop_{65-84}})\times 100$. The urban proxy II$_{77}$ is defined as $(1-\frac{PopBorn_{77}}{Totpop_{70-84}})\times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.9: Female years of schooling and cohort dummy$_{7678}$: Interaction of cohort dummy$_{7678}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{7678}$ × urban proxy I$_{7678}$ | -0.05$^{**}$ | -0.05$^{***}$ | -0.05$^{**}$ | -0.05$^{**}$ | -0.07$^{***}$ |
| | (0.017) | (0.016) | (0.016) | (0.016) | (0.015) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel B | | | | | |
| Cohort dummy$_{7678}$ × urban proxy II$_{7678}$ | -0.04$^{**}$ | -0.04$^{***}$ | -0.04$^{**}$ | -0.04$^{**}$ | -0.06$^{***}$ |
| | (0.014) | (0.013) | (0.013) | (0.013) | (0.013) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel C | | | | | |
| Cohort dummy$_{7678}$ × urban dummy | -0.25$^{***}$ | -0.25$^{***}$ | -0.24$^{***}$ | -0.24$^{***}$ | -0.50$^{***}$ |
| | (0.043) | (0.043) | (0.047) | (0.062) | (0.084) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel D | | | | | |
| Cohort dummy$_{7678}$ × big 3 urban dummy | -0.24$^{*}$ | -0.24$^{*}$ | -0.22 | -0.21 | -0.67$^{***}$ |
| | (0.133) | (0.132) | (0.138) | (0.134) | (0.177) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel E | | | | | |
| Cohort dummy$_{7678}$ × Phnom Penh dummy | -0.34 | -0.36 | -0.32 | -0.32 | -0.90$^{***}$ |
| | (0.240) | (0.231) | (0.232) | (0.229) | (0.246) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 198,477. The urban proxy I$_{7678}$ is constructed as $(1 - \frac{PopBorn_{7678}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7678}$ is defined as $(1 - \frac{PopBorn_{7678}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.10: Female years of schooling and cohort dummy$_{7579}$: Interaction of cohort dummy$_{7579}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{7579}$ × urban proxy I$_{7579}$ | -0.02 | -0.02 | -0.02 | -0.02 | -0.05*** |
| | (0.014) | (0.013) | (0.013) | (0.013) | (0.011) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel B | | | | | |
| Cohort dummy$_{7579}$ × urban proxy II$_{7579}$ | -0.02 | -0.02 | -0.02 | -0.02 | -0.04*** |
| | (0.011) | (0.011) | (0.011) | (0.011) | (0.009) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel C | | | | | |
| Cohort dummy$_{7579}$ × urban dummy | -0.12 | -0.13 | -0.11 | -0.11 | -0.44*** |
| | (0.095) | (0.095) | (0.099) | (0.101) | (0.118) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel D | | | | | |
| Cohort dummy$_{7579}$ × big 3 urban dummy | -0.11 | -0.12 | -0.09 | -0.09 | -0.67*** |
| | (0.114) | (0.112) | (0.117) | (0.119) | (0.129) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Panel E | | | | | |
| Cohort dummy$_{7579}$ × Phnom Penh dummy | -0.10 | -0.13 | -0.08 | -0.08 | -0.74*** |
| | (0.199) | (0.194) | (0.192) | (0.192) | (0.196) |
| $R^2$ | 0.15 | 0.15 | 0.16 | 0.16 | 0.17 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 198,477. The urban proxy I$_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.11: Male years of schooling and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.24*** | -0.22*** | -0.23*** | -0.22*** | -0.27*** |
| | (0.059) | (0.054) | (0.054) | (0.052) | (0.045) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel B | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.19*** | -0.18*** | -0.18*** | -0.18*** | -0.22*** |
| | (0.047) | (0.043) | (0.043) | (0.041) | (0.038) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel C | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.66*** | -0.72*** | -0.60*** | -0.65*** | -0.78*** |
| | (0.083) | (0.087) | (0.093) | (0.095) | (0.129) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel D | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.47*** | -0.52*** | -0.40** | -0.46** | -0.66** |
| | (0.160) | (0.153) | (0.163) | (0.161) | (0.284) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel E | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -1.38*** | -1.45*** | -1.24*** | -1.33*** | -1.71*** |
| | (0.135) | (0.104) | (0.109) | (0.101) | (0.127) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 185,567. The urban proxy I$_{77}$ is constructed as $(1-\frac{PopBorn_{77}}{Totpop_{65-84}})\times 100$. The urban proxy II$_{77}$ is defined as $(1-\frac{PopBorn_{77}}{Totpop_{70-84}})\times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.12: Male years of schooling and cohort $\text{dummy}_{7678}$: Interaction of cohort $\text{dummy}_{7678}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort $\text{dummy}_{7678}$ × urban proxy $\text{I}_{7678}$ | -0.06* | -0.06* | -0.05* | -0.05* | -0.08*** |
| | (0.032) | (0.028) | (0.030) | (0.028) | (0.021) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel B | | | | | |
| Cohort $\text{dummy}_{7678}$ × urban proxy $\text{II}_{7678}$ | -0.05* | -0.05** | -0.05* | -0.05* | -0.07*** |
| | (0.026) | (0.023) | (0.024) | (0.023) | (0.018) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel C | | | | | |
| Cohort $\text{dummy}_{7678}$ × urban dummy | -0.27* | -0.31** | -0.22 | -0.26 | -0.45** |
| | (0.147) | (0.150) | (0.150) | (0.153) | (0.171) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel D | | | | | |
| Cohort $\text{dummy}_{7678}$ × big 3 urban dummy | -0.38* | -0.38** | -0.29 | -0.31* | -0.62** |
| | (0.184) | (0.159) | (0.182) | (0.172) | (0.258) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel E | | | | | |
| Cohort $\text{dummy}_{7678}$ × Phnom Penh dummy | -0.65* | -0.64* | -0.52 | -0.54* | -1.02*** |
| | (0.347) | (0.305) | (0.313) | (0.300) | (0.288) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 185,567. The urban proxy $\text{I}_{7678}$ is constructed as $(1 - \frac{PopBorn_{7678}}{Totpop_{65-84}}) \times 100$. The urban proxy $\text{II}_{7678}$ is defined as $(1 - \frac{PopBorn_{7678}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
* indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table B.13: Male years of schooling and cohort $\text{dummy}_{7579}$: Interaction of cohort $\text{dummy}_{7579}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban proxy $\text{I}_{7579}$ | -0.02 | -0.02 | -0.02 | -0.02 | -0.04** |
| | (0.021) | (0.018) | (0.020) | (0.018) | (0.013) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel B | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban proxy $\text{II}_{7579}$ | -0.02 | -0.02 | -0.01 | -0.02 | -0.03*** |
| | (0.017) | (0.015) | (0.016) | (0.015) | (0.011) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel C | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban dummy | -0.11 | -0.16 | -0.05 | -0.10 | -0.35*** |
| | (0.108) | (0.113) | (0.118) | (0.122) | (0.121) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel D | | | | | |
| Cohort $\text{dummy}_{7579}$ × big 3 urban dummy | -0.12 | -0.16* | -0.02 | -0.07 | -0.44*** |
| | (0.118) | (0.081) | (0.117) | (0.102) | (0.153) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Panel E | | | | | |
| Cohort $\text{dummy}_{7579}$ × Phnom Penh dummy | -0.21 | -0.27* | -0.10 | -0.16 | -0.69*** |
| | (0.192) | (0.155) | (0.167) | (0.158) | (0.121) |
| $R^2$ | 0.16 | 0.18 | 0.18 | 0.19 | 0.20 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 185,567. The urban proxy $\text{I}_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy $\text{II}_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.14: Years of schooling and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables (Census 2008).

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.15*** | -0.16*** | -0.14*** | -0.15*** | -0.15*** | -0.14*** |
| | (0.019) | (0.019) | (0.023) | (0.018) | (0.021) | (0.024) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel B | | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.12*** | -0.12*** | -0.11*** | -0.12*** | -0.12*** | -0.11*** |
| | (0.014) | (0.014) | (0.017) | (0.014) | (0.016) | (0.019) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel C | | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.22*** | -0.19** | -0.20*** | -0.24*** | -0.20*** | -0.24*** |
| | (0.068) | (0.070) | (0.056) | (0.069) | (0.063) | (0.065) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel D | | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.33*** | -0.30*** | -0.33*** | -0.35*** | -0.32*** | -0.37*** |
| | (0.093) | (0.095) | (0.079) | (0.095) | (0.090) | (0.095) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel E | | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -0.70*** | -0.65*** | -0.63*** | -0.68*** | -0.62*** | -0.69*** |
| | (0.093) | (0.091) | (0.053) | (0.107) | (0.079) | (0.067) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 2008 Census survey. The total number of observations is 358,440. The urban proxy I$_{77}$ is constructed as $(1-\frac{PopBorn_{77}}{Totpop_{65-84}})\times 100$. The urban proxy II$_{77}$ is defined as $(1-\frac{PopBorn_{77}}{Totpop_{70-84}})\times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.15: Years of schooling and cohort $\text{dummy}_{7678}$: Interaction of cohort $\text{dummy}_{7678}$ with urban-born variables (Census 2008).

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort $\text{dummy}_{7678}$ × urban proxy $\text{I}_{7678}$ | -0.05*** | -0.05*** | -0.04*** | -0.05*** | -0.05*** | -0.05*** |
| | (0.012) | (0.012) | (0.013) | (0.012) | (0.012) | (0.014) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel B | | | | | | |
| Cohort $\text{dummy}_{7678}$ × urban proxy $\text{II}_{7678}$ | -0.04*** | -0.04*** | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| | (0.010) | (0.010) | (0.011) | (0.010) | (0.010) | (0.011) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel C | | | | | | |
| Cohort $\text{dummy}_{7678}$ × urban dummy | -0.15* | -0.16* | -0.13* | -0.17* | -0.17** | -0.26*** |
| | (0.083) | (0.082) | (0.067) | (0.083) | (0.073) | (0.076) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel D | | | | | | |
| Cohort $\text{dummy}_{7678}$ × big 3 urban dummy | -0.22 | -0.24 | -0.20 | -0.24 | -0.24* | -0.35** |
| | (0.146) | (0.152) | (0.119) | (0.139) | (0.132) | (0.141) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel E | | | | | | |
| Cohort $\text{dummy}_{7678}$ × Phnom Penh dummy | -0.74 | -0.72*** | -0.66*** | -0.71*** | -0.67*** | -0.83*** |
| | (.) | (0.076) | (0.052) | (0.111) | (0.085) | (0.050) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 2008 Census survey. The total number of observations is 358,440. The urban proxy $\text{I}_{7678}$ is constructed as $(1-\frac{PopBorn_{7678}}{Totpop_{65-84}})\times 100$. The urban proxy $\text{II}_{7678}$ is defined as $(1-\frac{PopBorn_{7678}}{Totpop_{70-84}})\times 100$.The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.16: Years of schooling and cohort $\text{dummy}_{7579}$: Interaction of cohort $\text{dummy}_{7579}$ with urban-born variables (Census 2008).

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban proxy $\text{I}_{7579}$ | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** |
| | (0.008) | (0.009) | (0.009) | (0.008) | (0.009) | (0.009) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel B | | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban proxy $\text{II}_{7579}$ | -0.02*** | -0.02*** | -0.02*** | -0.03*** | -0.03*** | -0.03*** |
| | (0.007) | (0.007) | (0.007) | (0.006) | (0.007) | (0.007) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel C | | | | | | |
| Cohort $\text{dummy}_{7579}$ × urban dummy | -0.11 | -0.13 | -0.10 | -0.13 | -0.13 | -0.31*** |
| | (0.099) | (0.095) | (0.081) | (0.098) | (0.084) | (0.057) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel D | | | | | | |
| Cohort $\text{dummy}_{7579}$ × big 3 urban dummy | -0.15 | -0.16 | -0.13 | -0.17 | -0.17 | -0.35*** |
| | (0.132) | (0.133) | (0.107) | (0.130) | (0.118) | (0.115) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Panel E | | | | | | |
| Cohort $\text{dummy}_{7579}$ × Phnom Penh dummy | -0.38** | -0.37** | -0.34*** | -0.38** | -0.35*** | -0.60*** |
| | (0.146) | (0.134) | (0.111) | (0.148) | (0.116) | (0.094) |
| $R^2$ | 0.14 | 0.19 | 0.15 | 0.16 | 0.21 | 0.21 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 2008 Census survey. The total number of observations is 358,440. The urban proxy $\text{I}_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy $\text{II}_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.17: Female wealth index score and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.08*** | -0.08*** | -0.07*** | -0.07*** | -0.08** |
| | (0.019) | (0.019) | (0.018) | (0.018) | (0.037) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel B | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.04*** | -0.04*** | -0.04*** | -0.04*** | -0.04* |
| | (0.011) | (0.011) | (0.011) | (0.011) | (0.023) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel C | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.09 | -0.11 | -0.12 | -0.13 | -0.17 |
| | (0.074) | (0.075) | (0.075) | (0.074) | (0.205) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel D | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.62*** | -0.59*** | -0.61*** | -0.58*** | -0.71*** |
| | (0.071) | (0.074) | (0.070) | (0.073) | (0.211) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel E | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -1.33*** | -1.39*** | -1.30*** | -1.32*** | -1.79*** |
| | (0.158) | (0.149) | (0.150) | (0.141) | (0.121) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 12,466. The urban proxy I$_{77}$ is constructed as $(1 - \frac{PopBorn_{77}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{77}$ is defined as $(1 - \frac{PopBorn_{77}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.18: Female wealth index score and cohort dummy$_{7678}$: Interaction of cohort dummy$_{7678}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{7678}$ × urban proxy I$_{7678}$ | -0.01 | -0.01 | -0.01 | -0.01 | -0.00 |
| | (0.012) | (0.013) | (0.012) | (0.013) | (0.023) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel B | | | | | |
| Cohort dummy$_{7678}$ × urban proxy II$_{7678}$ | -0.01 | -0.01 | -0.01 | -0.01 | 0.00 |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.015) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel C | | | | | |
| Cohort dummy$_{7678}$ × urban dummy | -0.23* | -0.25* | -0.22 | -0.24* | -0.43** |
| | (0.130) | (0.136) | (0.138) | (0.139) | (0.166) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel D | | | | | |
| Cohort dummy$_{7678}$ × big 3 urban dummy | -0.44** | -0.46** | -0.44** | -0.45** | -0.63** |
| | (0.199) | (0.211) | (0.209) | (0.209) | (0.292) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel E | | | | | |
| Cohort dummy$_{7678}$ × Phnom Penh dummy | -0.58 | -0.61 | -0.56 | -0.57 | -0.98 |
| | (0.500) | (0.523) | (0.514) | (0.519) | (0.654) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 12,466. The urban proxy I$_{7678}$ is constructed as $(1 - \frac{PopBorn_{7678}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7678}$ is defined as $(1 - \frac{PopBorn_{7678}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.19: Female wealth index score and cohort dummy$_{7579}$: Interaction of cohort dummy$_{7579}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{7579}$ × urban proxy I$_{7579}$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.009) | (0.008) | (0.009) | (0.008) | (0.018) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel B | | | | | |
| Cohort dummy$_{7579}$ × urban proxy II$_{7579}$ | -0.00 | -0.00 | -0.00 | -0.00 | 0.00 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.014) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel C | | | | | |
| Cohort dummy$_{7579}$ × urban dummy | -0.22** | -0.23** | -0.21** | -0.22** | -0.43* |
| | (0.092) | (0.097) | (0.098) | (0.100) | (0.213) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel D | | | | | |
| Cohort dummy$_{7579}$ × big 3 urban dummy | -0.29 | -0.31 | -0.28 | -0.29 | -0.47 |
| | (0.192) | (0.191) | (0.200) | (0.195) | (0.329) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Panel E | | | | | |
| Cohort dummy$_{7579}$ × Phnom Penh dummy | -0.48 | -0.46 | -0.43 | -0.42 | -0.71 |
| | (0.505) | (0.519) | (0.524) | (0.522) | (0.697) |
| $R^2$ | 0.19 | 0.21 | 0.21 | 0.22 | 0.26 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 12,466. The urban proxy I$_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^*$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.20: Male wealth index score and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.01 | -0.02 | -0.01 | -0.01 | -0.01 |
| | (0.016) | (0.016) | (0.014) | (0.015) | (0.011) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel B | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | 0.00 | -0.00 | 0.00 | 0.00 | -0.00 |
| | (0.008) | (0.008) | (0.007) | (0.008) | (0.006) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel C | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.21*** | -0.19*** | -0.16*** | -0.15*** | -0.12** |
| | (0.015) | (0.017) | (0.013) | (0.017) | (0.048) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel D | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.35*** | -0.32*** | -0.32*** | -0.31*** | -0.19** |
| | (0.014) | (0.019) | (0.005) | (0.019) | (0.080) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel E | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -0.40*** | -0.27** | -0.30*** | -0.19* | -0.10 |
| | (0.110) | (0.106) | (0.100) | (0.097) | (0.278) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 55,130. The urban proxy I$_{77}$ is constructed as $(1 - \frac{PopBorn_{77}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{77}$ is defined as $(1 - \frac{PopBorn_{77}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.21: Male wealth index score and cohort dummy$_{7678}$: Interaction of cohort dummy$_{7678}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{7678}$ × urban proxy I$_{7678}$ | -0.02*** | -0.02*** | -0.02*** | -0.02*** | -0.00 |
| | (0.004) | (0.005) | (0.004) | (0.005) | (0.007) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel B | | | | | |
| Cohort dummy$_{7678}$ × urban proxy II$_{7678}$ | -0.01*** | -0.01*** | -0.01** | -0.01** | -0.00 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel C | | | | | |
| Cohort dummy$_{7678}$ × urban dummy | -0.14** | -0.19*** | -0.11** | -0.15*** | -0.18*** |
| | (0.057) | (0.043) | (0.044) | (0.035) | (0.053) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel D | | | | | |
| Cohort dummy$_{7678}$ × big 3 urban dummy | -0.35*** | -0.39*** | -0.31*** | -0.35*** | -0.38*** |
| | (0.062) | (0.075) | (0.062) | (0.073) | (0.132) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel E | | | | | |
| Cohort dummy$_{7678}$ × Phnom Penh dummy | -0.46 | -0.38 | -0.39 | -0.33 | -0.44 |
| | (0.317) | (0.286) | (0.297) | (0.274) | (0.331) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 55,130. urban proxy I$_{7678}$ is constructed as $(1 - \frac{PopBorn_{7678}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7678}$ is defined as $(1 - \frac{PopBorn_{7678}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.22: Male wealth index score and cohort dummy$_{7579}$: Interaction of cohort dummy$_{7579}$ with urban-born variables.

| | Dependent variable is the wealth index score | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A | | | | | |
| Cohort dummy$_{7579}$ $\times$ urban proxy I$_{7579}$ | -0.01 | -0.01* | -0.01 | -0.01 | 0.01 |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.007) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel B | | | | | |
| Cohort dummy$_{7579}$ $\times$ urban proxy II$_{7579}$ | -0.00 | -0.00 | -0.00 | -0.00 | 0.01 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel C | | | | | |
| Cohort dummy$_{7579}$ $\times$ urban dummy | -0.06 | -0.10* | -0.04 | -0.07 | -0.04 |
| | (0.051) | (0.053) | (0.045) | (0.048) | (0.090) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel D | | | | | |
| Cohort dummy$_{7579}$ $\times$ big 3 urban dummy | -0.12 | -0.17 | -0.08 | -0.13 | 0.02 |
| | (0.168) | (0.154) | (0.174) | (0.162) | (0.220) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Panel E | | | | | |
| Cohort dummy$_{7579}$ $\times$ Phnom Penh dummy | 0.06 | -0.00 | 0.13 | 0.07 | 0.25 |
| | (0.430) | (0.362) | (0.430) | (0.372) | (0.418) |
| $R^2$ | 0.12 | 0.15 | 0.15 | 0.17 | 0.18 |
| Marital status | No | Yes | No | Yes | Yes |
| Number of children | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 1998 Census survey. The total number of observations is 55,130. The urban proxy I$_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.23: Wealth index score and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables (Census 2008).

| | Dependent variable is the wealth index score | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.009) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel B | | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.02*** | -0.02*** | -0.02*** | -0.02*** | -0.02*** | -0.02*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.007) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel C | | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.04*** | -0.04*** | -0.04*** | -0.05*** | -0.04*** | -0.02 |
| | (0.010) | (0.010) | (0.009) | (0.010) | (0.009) | (0.016) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel D | | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.05** | -0.04** | -0.07*** | -0.08*** | -0.09*** | -0.07* |
| | (0.018) | (0.017) | (0.018) | (0.016) | (0.023) | (0.037) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel E | | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -0.56*** | -0.55*** | -0.60*** | -0.58*** | -0.60*** | -0.57*** |
| | (0.042) | (0.041) | (0.039) | (0.035) | (0.032) | (0.055) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 2008 Census survey. The total number of observations is 134,541. The urban proxy I$_{77}$ is constructed as $(1 - \frac{PopBorn_{77}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{77}$ is defined as $(1 - \frac{PopBorn_{77}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration. $^*$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.24: Wealth index score and cohort dummy$_{7678}$: Interaction of cohort dummy$_{7678}$ with urban-born variables (Census 2008).

| | Dependent variable is the wealth index score | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{7678}$ × urban proxy I$_{7678}$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.004) | (0.004) | (0.005) | (0.004) | (0.005) | (0.006) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel B | | | | | | |
| Cohort dummy$_{7678}$ × urban proxy II$_{7678}$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel C | | | | | | |
| Cohort dummy$_{7678}$ × urban dummy | -0.01 | -0.01 | -0.01 | -0.02 | -0.02 | 0.01 |
| | (0.023) | (0.023) | (0.022) | (0.021) | (0.019) | (0.015) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel D | | | | | | |
| Cohort dummy$_{7678}$ × big 3 urban dummy | -0.06 | -0.06 | -0.06 | -0.07$^{**}$ | -0.08$^{*}$ | -0.07 |
| | (0.039) | (0.038) | (0.044) | (0.032) | (0.038) | (0.054) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel E | | | | | | |
| Cohort dummy$_{7678}$ × Phnom Penh dummy | -0.32 | -0.32$^{***}$ | -0.32$^{***}$ | -0.28$^{***}$ | -0.28$^{***}$ | -0.28$^{***}$ |
| | (.) | (0.011) | (0.024) | (0.012) | (0.022) | (0.086) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 2008 Census survey. The total number of observations is 134,541. The urban proxy I$_{7678}$ is constructed as $(1 - \frac{PopBorn_{7678}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7678}$ is defined as $(1 - \frac{PopBorn_{7678}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.25: Wealth index score and cohort dummy$_{7579}$: Interaction of cohort dummy$_{7579}$ with urban-born variables (Census 2008).

| | Dependent variable is the wealth index score | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{7579}$ × urban proxy I$_{7579}$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel B | | | | | | |
| Cohort dummy$_{7579}$ × urban proxy II$_{7579}$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel C | | | | | | |
| Cohort dummy$_{7579}$ × urban dummy | -0.02 | -0.02 | -0.02 | -0.02** | -0.02*** | -0.01 |
| | (0.013) | (0.013) | (0.011) | (0.009) | (0.006) | (0.020) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel D | | | | | | |
| Cohort dummy$_{7579}$ × big 3 urban dummy | -0.02 | -0.02 | -0.02 | -0.03 | -0.03 | 0.00 |
| | (0.027) | (0.027) | (0.029) | (0.024) | (0.027) | (0.063) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Panel E | | | | | | |
| Cohort dummy$_{7579}$ × Phnom Penh dummy | -0.13* | -0.13* | -0.13 | -0.11 | -0.10 | -0.07 |
| | (0.077) | (0.075) | (0.083) | (0.076) | (0.078) | (0.138) |
| $R^2$ | 0.18 | 0.18 | 0.20 | 0.22 | 0.23 | 0.23 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is a head of the household in the 2008 Census survey. The total number of observations is 134,541. The urban proxy I$_{7579}$ is constructed as $(1 - \frac{PopBorn_{7579}}{Totpop_{65-84}}) \times 100$. The urban proxy II$_{7579}$ is defined as $(1 - \frac{PopBorn_{7579}}{Totpop_{70-84}}) \times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place. See Appendix Table B.27 for illustration.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.26: Body height or weight and cohort dummy.

| Dependent variables: | Height | | | | Weight | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel A | | | | | | | | |
| Cohort dummy$_{77}$ | -0.34 | -0.36 | -0.33 | -0.62 | -0.37 | -0.38 | -0.32 | -0.53 |
| | (0.371) | (0.440) | (0.463) | (0.535) | (0.749) | (0.786) | (0.822) | (0.788) |
| $R^2$ | 0.07 | 0.08 | 0.08 | 0.11 | 0.04 | 0.06 | 0.07 | 0.12 |
| Panel B | | | | | | | | |
| Cohort dummy$_{7678}$ | -0.01 | -0.05 | -0.01 | -0.28 | -0.00 | -0.11 | -0.02 | -0.17 |
| | (0.164) | (0.211) | (0.168) | (0.259) | (0.282) | (0.248) | (0.251) | (0.331) |
| $R^2$ | 0.07 | 0.08 | 0.08 | 0.11 | 0.04 | 0.06 | 0.07 | 0.12 |
| Panel C | | | | | | | | |
| Cohort dummy$_{7579}$ | 0.23 | 0.19 | 0.26** | -0.04 | -0.24 | -0.33 | -0.24 | -0.71** |
| | (0.136) | (0.151) | (0.105) | (0.098) | (0.286) | (0.280) | (0.246) | (0.310) |
| $R^2$ | 0.07 | 0.08 | 0.08 | 0.11 | 0.04 | 0.06 | 0.07 | 0.12 |
| Marital status | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Number of children | No | No | Yes | Yes | No | No | Yes | Yes |
| Birth month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Province fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Province-specific cubic time trends | No | No | No | Yes | No | No | No | Yes |
| Number of observations | 3,054 | 3,054 | 3,054 | 3,054 | 3,057 | 3,057 | 3,057 | 3,057 |

*Notes:* Ordinary least square (OLS) regressions are estimated with standard errors in parentheses clustered by birth month and province of residence. The unit of observation is a woman in the 2000 Demographic and Health Survey (DHS) data.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.27: Years of schooling and cohort dummy$_{77}$: Interaction of cohort dummy$_{77}$ with urban-born variables.

| | Dependent variable is the year of schooling | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A | | | | | | |
| Cohort dummy$_{77}$ × urban proxy I$_{77}$ | -0.20*** | -0.19*** | -0.18*** | -0.19*** | -0.18*** | -0.23*** |
| | (0.044) | (0.044) | (0.042) | (0.041) | (0.041) | (0.036) |
| $R^2$ | 0.13955 | 0.17673 | 0.15145 | 0.15587 | 0.19115 | 0.19816 |
| Panel B | | | | | | |
| Cohort dummy$_{77}$ × urban proxy II$_{77}$ | -0.16*** | -0.15*** | -0.15*** | -0.15*** | -0.14*** | -0.19*** |
| | (0.035) | (0.034) | (0.033) | (0.032) | (0.032) | (0.030) |
| $R^2$ | 0.13955 | 0.17673 | 0.15145 | 0.15586 | 0.19115 | 0.19816 |
| Panel C | | | | | | |
| Cohort dummy$_{77}$ × urban dummy | -0.50*** | -0.46*** | -0.50*** | -0.46*** | -0.43*** | -0.57*** |
| | (0.062) | (0.066) | (0.065) | (0.070) | (0.074) | (0.088) |
| $R^2$ | 0.13954 | 0.17671 | 0.15145 | 0.15585 | 0.19114 | 0.19814 |
| Panel D | | | | | | |
| Cohort dummy$_{77}$ × big 3 urban dummy | -0.56*** | -0.52*** | -0.56*** | -0.50*** | -0.47*** | -0.74*** |
| | (0.104) | (0.104) | (0.104) | (0.109) | (0.111) | (0.176) |
| $R^2$ | 0.13952 | 0.17669 | 0.15142 | 0.15583 | 0.19112 | 0.19812 |
| Panel E | | | | | | |
| Cohort dummy$_{77}$ × Phnom Penh dummy | -1.27*** | -1.23*** | -1.29*** | -1.19*** | -1.18*** | -1.63*** |
| | (0.103) | (0.098) | (0.090) | (0.092) | (0.089) | (0.093) |
| $R^2$ | 0.13953 | 0.17671 | 0.15144 | 0.15585 | 0.19114 | 0.19815 |
| Female | No | Yes | No | No | Yes | Yes |
| Marital status | No | No | Yes | No | Yes | Yes |
| Number of children | No | No | No | Yes | Yes | Yes |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District-specific cubic time trends | No | No | No | No | No | Yes |

*Notes:* Ordinary least square regressions are estimated, with standard errors in parentheses clustered by birth years and district birthplaces. The unit of observation is an individual in the 1998 Census survey. The total number of observations is 384,044. The urban proxy I$_{77}$ is constructed as $(1-\frac{PopBorn_{77}}{Totpop_{65-84}})\times 100$. The urban proxy II$_{77}$ is defined as $(1-\frac{PopBorn_{77}}{Totpop_{70-84}})\times 100$. The urban dummy is an indicator for all urban district birthplaces. The big 3 urban dummy refers to a dummy variable for six urban districts located in the three major urban cities: Battambang, Siem Reap, and Phnom Penh. The Phnom Penh dummy indicates four urban district birthplaces within Phnom Penh. The $R^2$ values in each panel are the same up to two decimal places but differ slightly at the fifth decimal place.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table B.28: Thirty-one urban districts.

| District code | District name | Province |
|---|---|---|
| 106 | Krong Serei Saophoan | Banteay Mean Chey |
| 203 | Krong Battambang | Battambang |
| 305 | Krong Kampong Cham | Kampong Cham |
| 403 | Krong Kampong Chhnang | Kampong Chhnang |
| 502 | Krong Chbar Mon | Kampong Spueu |
| 603 | Krong Stueng Saen | Kampong Thum |
| 708 | Kampong Bay | Kampot |
| 811 | Krong Ta Khmau | Kandal |
| 904 | Smach Mean Chey | Kaoh Kong |
| 1002 | Kracheh | Kracheh |
| 1105 | Krong Saen Monourom | Mondol Kiri |
| 1201 | Chamkar Mon | Phnom Penh |
| 1202 | Doun Penh | |
| 1203 | Prampir Meakkakra | |
| 1204 | Tuol Kouk | |
| 1307 | Tbaeng Mean Chey | Preah Vihear |
| 1411 | Kampong Leav | Prey Veaeng |
| 1505 | Sampov Meas | Pousat |
| 1602 | Krong Ban Lung | Rotanak Kiri |
| 1710 | Krong Siem Reab | Siem Reap |
| 1801 | Krong Preah Sihanouk | Preah Sihanouk |
| 1802 | Prey Nob | |
| 1803 | Stueng Hav | |
| 1904 | Krong Stueng Traeng | Stueng Traeng |
| 2006 | Krong Svay Rieng | Svay Rieng |
| 2108 | Krong Doun Kaev | Takaev |
| 2204 | Krong Samraong | Otdar Mean Chey |
| 2301 | Damnak Chang'aeur | Kaeb |
| 2302 | Krong Kaeb | |
| 2401 | Krong Pailin | Pailin |
| 2402 | Sala Krau | |

*Notes:* District code refers to the code assigned to birthplace districts in the 1998 Census– IPUMS International. The definition of urban districts is based on information from the National Institute of Statistics, Ministry of Planning, as reported in the General Population Census of Cambodia 1998: Final Census Results.

Table B.29: Variable availability for constructing wealth index scores.

| Variable | | | Census data 1998 | Census data 2008 |
|---|---|---|---|---|
| Durable assets | | | | |
| | | Radio | | X |
| | | TV | | X |
| | | Phone | | X |
| | | Cell | | X |
| | | PC | | X |
| | | Bike | | X |
| | | Motobike | | X |
| | | Auto | | X |
| | | Boat | | X |
| | | Tractor | | X |
| | | Hand Tractor | | X |
| | | Internet access at home | | X |
| | | Internet access outside home | | X |
| Utilities | | | | |
| | Water source | | | |
| | | Piped water | X | X |
| | | Tube or piped well | X | X |
| | | Dug well | X | X |
| | | Spring or river | X | X |
| | | Bought | X | X |
| | | Other | X | X |
| | Toilet | | X | X |
| | Fuel used for cooking | | | |
| | | Firewood | X | X |
| | | Charcoal | X | X |
| | | Kerosene | X | X |
| | | Liquefied petroleum gas | X | X |
| | | Electricity | X | X |
| | | None | X | X |
| | | Other | X | X |
| | Fuel used for lighting | | | |
| | | City power | X | X |
| | | Generator | X | X |
| | | City and Generator | X | X |
| | | Kerosene | X | X |
| | | Candle | X | X |
| | | Battery | X | X |
| | | Other | X | X |
| Dwelling characteristics | | | | |
| | Member per room or bedroom | | X | X |
| Total Variables | | | 22 | 35 |

*Notes:* The data source is from the Minnesota Population Center's Integrated Public Use Microdata Series - International (IPUMS-I). An X indicates the availability of a variable in the census data.

# Appendix C

## Chapter 3 Appendix

Table C.1: Fraction of top ten crops grown on farms with hired workers.

| | Mean | Std. Error | Min. | Max. | Median | Obs. |
|---|---|---|---|---|---|---|
| Fraction rice | 0.82 | 0.35 | 0 | 1 | 1.00 | 5,132 |
| Fraction cassava | 0.07 | 0.21 | 0 | 1 | 0.00 | 5,132 |
| Fraction cashew | 0.03 | 0.14 | 0 | 1 | 0.00 | 5,132 |
| Fraction maize | 0.02 | 0.12 | 0 | 1 | 0.00 | 5,132 |
| Fraction mango | 0.02 | 0.11 | 0 | 1 | 0.00 | 5,132 |
| Fruit-bearing vegetables | 0.01 | 0.08 | 0 | 1 | 0.00 | 5,132 |
| Fraction rubber | 0.01 | 0.09 | 0 | 1 | 0.00 | 5,132 |
| Vegetables tree/leaf/flower | 0.00 | 0.05 | 0 | 1 | 0.00 | 5,132 |
| Oil and fiber | 0.00 | 0.04 | 0 | 1 | 0.00 | 5,132 |
| Sugarcane | 0.00 | 0.03 | 0 | 1 | 0.00 | 5,132 |
| Fraction other thirty-two crops | 0.02 | 0.11 | 0 | 1 | 0.00 | 5,132 |

*Notes:* The Cambodia Inter-Censal Agriculture Survey (CIAS) 2019 collected data on 42 different crops. The dataset includes 5,132 farm households that employed occasional and external workers across provinces.

Table C.2: Summary statistics of main variables by rice dummy.

| | Mean | Std. Error | Min. | Max. | Median | Obs. |
|---|---|---|---|---|---|---|
| A. Farms where the fraction of rice is greater than zero. | | | | | | |
| Fraction female workers | 0.17 | 0.29 | 0.00 | 1.00 | 0.00 | 4,528 |
| Fraction rice | 0.93 | 0.19 | 0.02 | 1.00 | 1.00 | 4,528 |
| Total farm size | 2.58 | 4.86 | 0.00 | 124.06 | 1.20 | 4,528 |
| Log total farm size | 0.28 | 1.11 | -7.60 | 4.82 | 0.18 | 4,528 |
| Total number of workers | 6.92 | 11.52 | 1.00 | 194.00 | 4.00 | 4,528 |
| Log total number of workers | 1.44 | 0.86 | 0.00 | 5.27 | 1.39 | 4,528 |
| Average family schooling | 4.59 | 2.65 | 0.00 | 18.00 | 4.25 | 4,528 |
| Fraction female in households | 0.31 | 0.20 | 0.00 | 1.00 | 0.25 | 4,528 |
| B. Farms where the fraction of rice equals zero. | | | | | | |
| Fraction female workers | 0.47 | 0.25 | 0.00 | 1.00 | 0.50 | 604 |
| Fraction rice | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 604 |
| Total farm size | 8.88 | 19.89 | 0.00 | 330.00 | 3.50 | 604 |
| Log total farm size | 1.16 | 1.59 | -7.60 | 5.80 | 1.25 | 604 |
| Total number of workers | 13.30 | 12.80 | 1.00 | 105.00 | 10.00 | 604 |
| Log total number of workers | 2.16 | 0.97 | 0.00 | 4.65 | 2.30 | 604 |
| Average family schooling | 5.13 | 2.92 | 0.00 | 18.00 | 4.67 | 604 |
| Fraction female in households | 0.26 | 0.17 | 0.00 | 1.00 | 0.25 | 604 |

*Notes:* The Cambodia Inter-Censal Agriculture Survey (CIAS) 2019 collected data on 42 different crops. Farms were categorized into two types: those used for rice paddy cultivation and those for non-rice paddy crops. The fraction rice is calculated as the total land area used for rice paddy divided by the total farm size, while the fraction non-rice paddy is computed as one minus the fraction rice. The dataset includes 5,132 farm households that employed occasional and external workers across provinces. The fraction of female workers refers to the proportion of occasional and external workers who are female. Total farm size is measured in hectares.

Table C.3: Summary statistics of district-level main variables.

| | Mean | Std. Error | Min. | Max. | Median | Obs. |
|---|---|---|---|---|---|---|
| Fraction female among agricultural occupation | 0.50 | 0.04 | 0.29 | 0.60 | 0.51 | 190 |
| Log actual rice yield in 2010 | 0.97 | 0.09 | 0.59 | 1.18 | 1.00 | 187 |
| Fraction female in workforce | 0.52 | 0.02 | 0.48 | 0.57 | 0.52 | 190 |
| Fraction population in agriculture | 0.29 | 0.15 | 0.00 | 0.53 | 0.32 | 190 |
| Log population density | 4.88 | 1.51 | 1.54 | 10.40 | 4.90 | 190 |
| Log precipitation | 4.87 | 0.23 | 4.45 | 5.78 | 4.81 | 190 |
| Log temperature | 3.31 | 0.03 | 3.20 | 3.35 | 3.32 | 190 |
| Log caloric suitability index | 7.00 | 0.06 | 6.89 | 7.15 | 6.98 | 190 |
| Log distance from district to Phnom Penh | 4.64 | 1.05 | 1.23 | 5.93 | 4.84 | 190 |
| Log distance from district to river | 4.27 | 0.75 | 1.13 | 5.51 | 4.29 | 190 |
| Log elevation | 3.40 | 1.20 | 1.11 | 6.21 | 3.24 | 190 |

Table C.4: Fraction females and rice: Non-rice as base.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Fraction rice | -0.32*** | -0.31*** | -0.19*** | -0.19*** | -0.19*** | -0.13*** | -0.13*** |
| | (0.010) | (0.011) | (0.012) | (0.012) | (0.012) | (0.013) | (0.044) |
| Log total farm size | | 0.01 | -0.04*** | -0.04*** | -0.04*** | -0.04*** | -0.04*** |
| | | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.009) |
| Log total number of workers | | | 0.18*** | 0.18*** | 0.18*** | 0.16*** | 0.16*** |
| | | | (0.004) | (0.004) | (0.004) | (0.005) | (0.017) |
| Average family schooling | | | | -0.00*** | -0.00*** | -0.00 | -0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | -0.01 | 0.02 | 0.02 |
| | | | | | (0.018) | (0.017) | (0.022) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.16 | 0.16 | 0.41 | 0.41 | 0.41 | 0.50 | 0.50 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary Least Squares regressions are weighted by the log of one plus total number of workers. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are categorized into two types: those used for rice paddy, and those used for non-rice paddy. Fraction rice is defined as the total land used for rice paddy divided by the total farm size. Fraction non-rice paddy equals one minus fraction rice.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.5: Fraction females and rice: Non-rice as base.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Fraction rice | -0.36*** | -0.35*** | -0.25*** | -0.25*** | -0.25*** | -0.18*** | -0.18*** |
| | (0.011) | (0.012) | (0.013) | (0.013) | (0.013) | (0.014) | (0.043) |
| Log total farm size | | 0.00 | -0.04*** | -0.03*** | -0.03*** | -0.04*** | -0.04*** |
| | | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) | (0.009) |
| Log total number of workers | | | 0.14*** | 0.14*** | 0.14*** | 0.14*** | 0.14*** |
| | | | (0.005) | (0.005) | (0.005) | (0.005) | (0.011) |
| Average family schooling | | | | -0.00*** | -0.00*** | 0.00 | 0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | 0.01 | 0.04** | 0.04* |
| | | | | | (0.021) | (0.019) | (0.022) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.23 | 0.23 | 0.41 | 0.41 | 0.41 | 0.51 | 0.51 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary Least Squares regressions are weighted by the log of one plus total farm size. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are categorized into two types: those used for rice paddy, and those used for non-rice paddy. Fraction rice is defined as the total land used for rice paddy divided by the total farm size. Fraction non-rice paddy equals one minus fraction rice.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.6: Fraction females and rice dummy.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Rice dummy | -0.30*** | -0.28*** | -0.19*** | -0.19*** | -0.19*** | -0.14*** | -0.14*** |
| | (0.011) | (0.012) | (0.013) | (0.013) | (0.013) | (0.013) | (0.034) |
| Log total farm size | | 0.02*** | -0.02*** | -0.02*** | -0.02*** | -0.03*** | -0.03*** |
| | | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) | (0.008) |
| Log total number of workers | | | 0.17*** | 0.17*** | 0.17*** | 0.16*** | 0.16*** |
| | | | (0.004) | (0.004) | (0.004) | (0.004) | (0.016) |
| Average family schooling | | | | -0.00*** | -0.00*** | 0.00 | 0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | -0.01 | 0.03 | 0.03 |
| | | | | | (0.018) | (0.017) | (0.022) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.11 | 0.11 | 0.35 | 0.35 | 0.35 | 0.45 | 0.45 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary least square (OLS) regressions. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are categorized into two types: those used for rice paddy, and those used for non-rice paddy. Fraction rice is defined as the total land used for rice paddy divided by the total farm size. Fraction non-rice paddy equals one minus fraction rice. The rice dummy is an indicator for fraction rice, equal to one when fraction rice is greater than zero, and zero otherwise.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.7: Fraction females and top ten crops: Other as base.

| | Dependent variable is the fraction female workers | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Fraction rice | -0.16*** | | | | | | | | | |
| | (0.040) | | | | | | | | | |
| Fraction cassava | | 0.09* | | | | | | | | |
| | | (0.046) | | | | | | | | |
| Fraction cashew | | | 0.11** | | | | | | | |
| | | | (0.041) | | | | | | | |
| Fraction maize | | | | 0.14*** | | | | | | |
| | | | | (0.038) | | | | | | |
| Fraction mango | | | | | 0.01 | | | | | |
| | | | | | (0.043) | | | | | |
| Fruit-bearing vegetables | | | | | | 0.27*** | | | | |
| | | | | | | (0.058) | | | | |
| Fraction rubber | | | | | | | 0.13*** | | | |
| | | | | | | | (0.032) | | | |
| Vegetables tree/leaf/flower | | | | | | | | 0.39*** | | |
| | | | | | | | | (0.072) | | |
| Oil and fiber | | | | | | | | | -0.01 | |
| | | | | | | | | | (0.038) | |
| Sugarcane | | | | | | | | | | 0.05 |
| | | | | | | | | | | (0.134) |
| Province fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard errors | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster | Cluster |
| $R^2$ | 0.46 | 0.44 | 0.44 | 0.44 | 0.44 | 0.44 | 0.44 | 0.44 | 0.44 | 0.44 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary least square (OLS) regressions. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are grouped into two types: those used for rice paddy and non-rice, with similar groupings for the other top nine crop types. Fraction rice is defined as the total land used for rice paddy divided by the total farm size, and fraction non-rice equals one minus fraction rice. The same calculation applies to the other top nine crops, with similar definitions for each. All control variables are included in the regressions but are not reported in this table. These control variables include Log total farm size, Log total number of workers, Average family schooling, and Fraction female in households.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.8: Fraction females and rice: Maize as base.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Fraction rice | -0.28*** | -0.28*** | -0.11*** | -0.11*** | -0.11*** | -0.13*** | -0.13*** |
| | (0.019) | (0.020) | (0.018) | (0.018) | (0.018) | (0.021) | (0.031) |
| Log total farm size | | 0.01* | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** |
| | | (0.004) | (0.003) | (0.003) | (0.003) | (0.004) | (0.009) |
| Log total number of workers | | | 0.19*** | 0.19*** | 0.19*** | 0.17*** | 0.17*** |
| | | | (0.004) | (0.004) | (0.004) | (0.004) | (0.018) |
| Average family schooling | | | | -0.00*** | -0.00*** | 0.00 | 0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | -0.01 | 0.03* | 0.03 |
| | | | | | (0.018) | (0.017) | (0.020) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.03 | 0.03 | 0.33 | 0.33 | 0.33 | 0.46 | 0.46 |
| Number of observations | 4,662 | 4,662 | 4,662 | 4,662 | 4,662 | 4,662 | 4,662 |

*Notes:* Ordinary least square (OLS) regressions. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are categorized into two types: those used for rice paddy, and those used for maize. Fraction rice is defined as the total land used for rice paddy divided by the total farm size. Fraction maize equals one minus fraction rice.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.9: Fraction females and forty-two crops: Fruit-bearing vegetables as base.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Fraction rice | -0.39*** | -0.39*** | -0.35*** | -0.36*** | -0.36*** | -0.30*** | -0.30*** |
| | (0.054) | (0.055) | (0.054) | (0.054) | (0.054) | (0.055) | (0.062) |
| Log total farm size | | 0.01* | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** |
| | | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) | (0.009) |
| Log total number of workers | | | 0.17*** | 0.17*** | 0.17*** | 0.16*** | 0.16*** |
| | | | (0.004) | (0.004) | (0.004) | (0.004) | (0.017) |
| Average family schooling | | | | -0.00*** | -0.00*** | 0.00 | 0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | -0.00 | 0.03* | 0.03 |
| | | | | | (0.018) | (0.017) | (0.022) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.18 | 0.18 | 0.38 | 0.38 | 0.38 | 0.47 | 0.47 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary least square (OLS) regressions. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are grouped into forty-two types based on the main crop cultivated (e.g., rice paddy, maize, and so on). Fraction rice is defined as the total land used for rice paddy divided by the total farm size, with similar definitions for fraction maize and the fraction of each remaining crop type.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.10: Fraction females, and rice and maize: Other crops as base.

| | Dependent variable is the fraction female workers | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Fraction rice | -0.36*** | -0.36*** | -0.25*** | -0.25*** | -0.25*** | -0.16*** | -0.16*** |
| | (0.011) | (0.013) | (0.013) | (0.013) | (0.013) | (0.015) | (0.042) |
| Fraction maize | -0.08** | -0.08** | -0.11*** | -0.11*** | -0.11*** | 0.01 | 0.01 |
| | (0.033) | (0.033) | (0.030) | (0.030) | (0.030) | (0.031) | (0.032) |
| Log total farm size | | 0.01 | -0.03*** | -0.03*** | -0.03*** | -0.03*** | -0.03*** |
| | | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) | (0.008) |
| Log total number of workers | | | 0.16*** | 0.16*** | 0.16*** | 0.16*** | 0.16*** |
| | | | (0.004) | (0.004) | (0.004) | (0.004) | (0.017) |
| Average family schooling | | | | -0.00*** | -0.00*** | 0.00 | 0.00 |
| | | | | (0.001) | (0.001) | (0.001) | (0.002) |
| Fraction female in households | | | | | -0.00 | 0.03* | 0.03 |
| | | | | | (0.018) | (0.017) | (0.022) |
| Province fixed effects | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.17 | 0.17 | 0.37 | 0.37 | 0.37 | 0.46 | 0.46 |
| Number of observations | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 | 5,132 |

*Notes:* Ordinary least square (OLS) regressions. Robust standard errors are in parentheses. For columns 7, standard errors are clustered by province. The unit of observation is farm households that employ occasional and external workers across provinces. Farms are categorized into three types: those used for rice paddy, maize, and other crops. Fraction rice is the total land used for rice paddy divided by the total farm size, and similarly for fraction maize. Fraction other is equal to one minus the sum of fraction rice and fraction maize.

$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.

Table C.11: Fraction females among agricultural occupation and log actual rice yield in 2010.

| | Dependent variable is the fraction female among agricultural occupation | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| Log actual rice yield in 2010 | -0.02 | -0.05* | -0.03 | -0.02 | -0.04* | -0.04** | -0.04** | -0.04** | -0.03 | -0.04** | -0.04* |
| | (0.028) | (0.025) | (0.020) | (0.019) | (0.018) | (0.017) | (0.018) | (0.018) | (0.020) | (0.021) | (0.025) |
| Fraction female in workforce | | 0.83*** | 0.72*** | 1.02*** | 1.09*** | 1.00*** | 0.98*** | 1.00*** | 1.01*** | 0.75*** | 0.75** |
| | | (0.158) | (0.134) | (0.209) | (0.203) | (0.186) | (0.195) | (0.202) | (0.196) | (0.229) | (0.347) |
| Fraction population in agriculture | | | 0.16*** | 0.12*** | 0.15*** | 0.14*** | 0.14*** | 0.14*** | 0.14*** | 0.11*** | 0.11*** |
| | | | (0.025) | (0.027) | (0.026) | (0.026) | (0.026) | (0.026) | (0.025) | (0.034) | (0.030) |
| Log population density | | | | -0.01* | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | | | | (0.004) | (0.004) | (0.005) | (0.005) | (0.006) | (0.006) | (0.007) | (0.004) |
| Log precipitation | | | | | 0.05** | 0.06** | 0.05* | 0.05* | 0.04 | -0.00 | -0.00 |
| | | | | | (0.020) | (0.026) | (0.030) | (0.030) | (0.030) | (0.036) | (0.049) |
| Log temperature | | | | | | 0.24 | 0.22 | 0.21 | 0.21 | 0.10 | 0.10 |
| | | | | | | (0.212) | (0.224) | (0.224) | (0.220) | (0.264) | (0.356) |
| Log caloric suitability index | | | | | | | -0.03 | -0.05 | -0.03 | -0.01 | -0.01 |
| | | | | | | | (0.071) | (0.070) | (0.071) | (0.110) | (0.125) |
| Log distance from district to Phnom Penh | | | | | | | | 0.01 | 0.01* | 0.01 | 0.01 |
| | | | | | | | | (0.005) | (0.005) | (0.011) | (0.013) |
| Log distance from district to river | | | | | | | | | 0.01** | 0.01 | 0.01 |
| | | | | | | | | | (0.003) | (0.005) | (0.004) |
| Province fixed effects | No | No | No | No | No | No | No | No | No | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Robust | Robust | Robust | Robust | Robust | Robust | Cluster |
| $R^2$ | 0.00 | 0.11 | 0.39 | 0.42 | 0.46 | 0.47 | 0.47 | 0.48 | 0.50 | 0.60 | 0.60 |
| Number of observations | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 |

*Notes:* Ordinary Least Squares (OLS) Regressions. Robust standard errors are in parentheses. In columns 1 through 10, standard errors are robust, while in column 11, they are clustered by province. The unit of observation is the district.
* indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table C.12: First-stage regression: Log actual rice yield in 2010 and log elevation.

| | Dependent variable is the log actual rice yield in 2010 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log elevation | -0.06*** | -0.08*** | -0.07*** | -0.07*** | -0.06*** | -0.08*** | -0.07*** | -0.07*** |
| | (0.010) | (0.012) | (0.013) | (0.013) | (0.010) | (0.012) | (0.013) | (0.013) |
| Fraction female in workforce | -0.32 | 0.09 | 0.10 | 0.14 | -0.32 | 0.09 | 0.10 | 0.14 |
| | (0.473) | (0.477) | (0.481) | (0.481) | (0.628) | (0.576) | (0.575) | (0.584) |
| Fraction population in agriculture | 0.03 | 0.07 | 0.06 | 0.05 | 0.03 | 0.07 | 0.06 | 0.05 |
| | (0.070) | (0.061) | (0.060) | (0.062) | (0.080) | (0.061) | (0.061) | (0.068) |
| Log population density | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.012) | (0.011) | (0.010) | (0.010) |
| Log precipitation | | 0.29*** | 0.27*** | 0.26*** | | 0.29*** | 0.27** | 0.26** |
| | | (0.081) | (0.083) | (0.082) | | (0.107) | (0.110) | (0.107) |
| Log temperature | | 0.36 | 0.42 | 0.25 | | 0.36 | 0.42 | 0.25 |
| | | (0.708) | (0.720) | (0.712) | | (0.830) | (0.859) | (0.832) |
| Log caloric suitability index | | 0.29 | 0.29 | 0.27 | | 0.29 | 0.29 | 0.27 |
| | | (0.346) | (0.347) | (0.339) | | (0.462) | (0.460) | (0.442) |
| Log distance from district to Phnom Penh | | | 0.02 | 0.03 | | | 0.02 | 0.03 |
| | | | (0.021) | (0.022) | | | (0.024) | (0.027) |
| Log distance from district to river | | | | -0.02 | | | | -0.02 |
| | | | | (0.012) | | | | (0.014) |
| Province fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard errors | Robust | Robust | Robust | Robust | Cluster | Cluster | Cluster | Cluster |
| Cragg-Donald Wald F statistic | 40.38 | 44.61 | 38.69 | 39.20 | 40.38 | 44.61 | 38.69 | 39.20 |
| Stock-Wright LM S statistic | 5.48 | 11.60 | 15.55 | 16.49 | 2.90 | 5.30 | 9.65 | 9.85 |
| First stage F-statistics | 9.47 | 9.28 | 8.46 | 8.02 | 7.89 | 14.12 | 14.44 | 14.84 |
| $R^2$ | 0.46 | 0.52 | 0.52 | 0.53 | 0.46 | 0.52 | 0.52 | 0.53 |
| Adjusted $R^2$ | 0.37 | 0.42 | 0.42 | 0.43 | 0.37 | 0.42 | 0.42 | 0.43 |
| Number of observations | 187 | 187 | 187 | 187 | 187 | 187 | 187 | 187 |

*Notes:* First Stage of Two-Stage Least Squares (2SLS) Regressions. Robust standard errors are in parentheses. In columns 5 through 8, standard errors are clustered by province. The unit of observation is the district.
$^{*}$ indicates $p < 0.10$, $^{**}$ $p < 0.05$, and $^{***}$ $p < 0.01$.