

# Cocoa Booms and Human Capital in Ghana: Evidence from Census Microdata

APEP Autonomous Research\*

March 10, 2026

## Abstract

World cocoa prices tripled between 2000 and 2010, delivering a massive income shock to Ghana—the world’s second-largest producer, where 800,000 farming households depend on the crop. Using 5.7 million individual records from three rounds of Ghana’s population census (1984, 2000, 2010), we compare human capital and employment trajectories in high- versus low-cocoa-intensity regions within the ecologically similar forest belt. A Bartik-style specification and a doubly-robust difference-in-differences estimator yield consistent findings: adult employment fell by 6.8 percentage points per unit of cocoa share, confirmed by exact randomization inference over all 720 regional permutations (RI  $p = 0.014$ ). Literacy rose by 5.5 percentage points under conventional inference, though exact permutation tests (RI  $p = 0.34$ ) underscore the challenge of six clusters. Pre-trends using 1984–2000—when cocoa prices were falling—show precisely zero differential trends.

**JEL Codes:** I25, O12, O13, Q17

**Keywords:** cocoa, commodity boom, human capital, education, child labor, Ghana, doubly-robust

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: N/A).

## 1. Introduction

In October 2000, a metric ton of cocoa sold for \$906 on world markets. By 2010, that same ton fetched \$2,800. For the 800,000 farming households that grow Ghana’s cocoa—the country’s largest export and 30% of its foreign exchange—this 209% price increase was a first-order change in economic conditions (Kolavalli and Vigneri, 2011). Whether such windfalls build or erode human capital is among the most consequential questions in development economics, and the answer is far from settled.

The theoretical tension is sharp. Higher commodity prices raise household income, relaxing the credit constraints that keep children out of school (Becker, 1965). But they simultaneously raise the returns to agricultural labor, pulling children into the fields (Kruger, 2007). In Ghana’s cocoa sector—where over 700,000 children were engaged in hazardous work as recently as 2020 (International Labour Organization and UNICEF, 2021)—the question of which force dominates is not academic.

We study this question using 5.7 million individual records from three rounds of Ghana’s Population and Housing Census (1984, 2000, 2010), accessed through IPUMS International. Our identification strategy exploits the sharp geographic concentration of cocoa production: four regions—Western, Ashanti, Brong Ahafo, and Eastern—account for over 95% of national output (Vigneri, 2005). We compare changes in education and employment outcomes between high- and low-cocoa-intensity regions within Ghana’s ecologically similar forest belt, using a Bartik-style specification that interacts pre-treatment regional cocoa shares with the world price shock. A doubly-robust difference-in-differences estimator (Sant’Anna and Zhao, 2020) provides a complementary approach that is consistent if either the outcome model or the propensity score model is correctly specified—a valuable property given the observable differences between cocoa and non-cocoa regions.

Our most robust finding is that the boom *reduced* adult employment. Employment rates fell by 6.8 percentage points per unit of cocoa share, confirmed by exact randomization inference over all 720 permutations of regional treatment assignments (RI  $p = 0.014$ ) and stable across every leave-one-region-out specification. The DR DiD confirms a 3.1 percentage point decline ( $p < 0.001$ ). Rather than intensifying agricultural labor, the windfall financed an exit from it—consistent with the structural transformation models of Lewis (1954) and McMillan et al. (2014).

The boom also improved educational outcomes for children aged 6–17, though precision depends on the inference approach. Literacy rose by 5.5 percentage points per unit of cocoa share under conventional clustered inference ( $p < 0.05$ ), and the DR DiD estimates a 3.1 percentage point gain ( $p < 0.001$ ). But exact permutation tests yield an RI  $p$ -value of 0.34,

illustrating the well-known challenge of inference with six clusters (Cameron et al., 2008). The literacy coefficient is positive in all six leave-one-region-out specifications (range: 0.041–0.103), suggesting the sign is reliable even if the conventional significance level overstates certainty.

Pre-trend tests strongly support identification. Using the 1984–2000 period—when cocoa prices were *falling*—we estimate precisely zero differential trends: enrollment ( $\beta = 0.0005$ ,  $p = 0.98$ ), primary completion ( $\beta = -0.0001$ ,  $p = 0.997$ ), employment ( $\beta = 0.009$ ,  $p = 0.60$ ), and agricultural employment ( $\beta = 0.018$ ,  $p = 0.86$ ). The three-period event study confirms the pattern: flat before 2000, sharp effects only during the boom.

This paper contributes to the literature on commodity shocks and human capital in developing countries (Kruger, 2007; Edmonds and Pavcnik, 2005; Cogneau and Jedwab, 2012; Beegle et al., 2006; Shah and Steinberg, 2017). Our setting—smallholder cocoa farming with high poverty and significant school fees—is precisely where theory predicts income effects should dominate, and they do. The tree-crop production structure of cocoa, which limits short-run labor intensification, weakens the substitution channel that drives the negative results in Kruger (2007) (Brazilian coffee) and Black et al. (2005) (Appalachian coal). We also speak to the resource curse debate (van der Ploeg, 2011; Sachs and Warner, 2001): when commodity windfalls flow to smallholder households rather than governments or corporations, they can finance human capital accumulation that outlasts the boom itself.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes cocoa production in Ghana and the 2000s price boom. Section 4 presents a conceptual framework. Section 5 describes our data. Section 6 details the empirical strategy. Section 7 presents results. Section 8 discusses robustness. Section 10 concludes.

## 2. Related Literature

Our paper sits at the intersection of four literatures.

### 2.1 Commodity Prices and Human Capital

Commodity price swings force a choice between the classroom and the field, and the evidence on which wins is strikingly mixed. Kruger (2007) provides the canonical finding that substitution effects dominate: coffee booms in Brazil reduced enrollment among children aged 14–17, who were most productive in harvesting. Cogneau and Jedwab (2012) study the opposite shock—the 1990 cocoa price collapse in Côte d’Ivoire—and find that the crisis reduced schooling, suggesting income effects matter when prices fall. The asymmetry hints at heterogeneity across crops, production structures, and initial conditions.

In Vietnam, [Edmonds and Pavcnik \(2005\)](#) documents that rice price liberalization raised enrollment for children under 15—income effects dominating in a smallholder setting. [Edmonds \(2005\)](#) provides complementary evidence that child labor declines as economic status improves. [Beegle et al. \(2006\)](#) exploits agricultural shocks in Tanzania to show the symmetric result: negative income shocks increase child labor and reduce schooling.

Beyond agriculture, [Black et al. \(2005\)](#) find that the Appalachian coal boom reduced educational attainment by raising opportunity costs. [Charles et al. \(2018\)](#) document the same for U.S. housing booms, and [Atkin \(2016\)](#) for export manufacturing in Mexico. These middle- and high-income settings consistently show substitution effects dominating—suggesting that income effects may be more important where credit constraints on schooling are severe.

Ghana’s cocoa sector has exactly these features: smallholder production, severe credit constraints, and a tree-crop technology that limits short-run labor intensification. We find that income effects dominate, consistent with the pattern [Edmonds \(2008\)](#) identifies: child labor is primarily a poverty phenomenon.

## 2.2 Income Shocks, Schooling, and Child Labor

Beyond commodity prices, a broader literature links household income to children’s schooling. [Maccini and Yang \(2009\)](#) show that early-life rainfall shocks in Indonesia have persistent effects on women’s educational attainment. [Shah and Steinberg \(2017\)](#) provides causal evidence that Indian droughts reduce human capital both contemporaneously and in the long run, concentrated among the poorest households. [Duflo \(2001\)](#) demonstrates that school construction in Indonesia had large effects on attainment and earnings, establishing the sensitivity of schooling decisions to economic conditions in developing countries.

The theoretical foundation comes from [Basu and Van \(1998\)](#), who model child labor as a consequence of poverty: families send children to work only when adult wages fall below a subsistence threshold. Their “luxury axiom”—that child labor is a bad that households consume less of as they grow richer—predicts that positive income shocks should reduce child labor and increase schooling. Our results are consistent with this prediction.

## 2.3 Resource Booms and Development

The “resource curse” hypothesis—that natural resource windfalls hinder development through Dutch disease, rent-seeking, and institutional deterioration ([Sachs and Warner, 2001](#); [van der Ploeg, 2011](#))—has been challenged by recent microeconomic evidence. [Allcott and Keniston \(2018\)](#) find that U.S. oil booms raise local wages with no manufacturing decline. [Caselli and Michaels \(2013\)](#) show that Brazilian oil windfalls improve local living standards. [Dube](#)

and Vargas (2013) find heterogeneous effects of commodity prices on conflict in Colombia, suggesting the structure of the commodity sector matters critically.

Our paper examines human capital—a channel through which resource booms could affect long-run development that the resource curse literature has largely overlooked. If commodity windfalls increase schooling, they lay the foundation for sustained growth even after prices decline. The structural transformation we document—workers exiting agriculture—is consistent with McMillan et al. (2014)’s evidence that structural change in Africa since 2000 has been growth-enhancing, and with Gollin et al. (2016)’s analysis of commodity-driven urbanization.

## 2.4 Econometric Methods

We employ the doubly-robust DiD estimator of Sant’Anna and Zhao (2020), which is consistent if either the outcome model or the propensity score model is correctly specified. Our Bartik-style specification follows Goldsmith-Pinkham et al. (2020), with inference foundations from Adão et al. (2019) and Borusyak et al. (2022). The “shift” (world cocoa price change) is plausibly exogenous to any single Ghanaian region; the “shares” (regional cocoa production) reflect pre-treatment geographic specialization. We follow Roth (2022)’s caution about pre-test power and complement pre-trends with the sensitivity analysis of Cinelli and Hazlett (2020).

With only 6 forest-belt clusters, conventional cluster-robust standard errors may over-reject (Cameron et al., 2008; MacKinnon and Webb, 2017). We address this through three complementary approaches: DR DiD influence-function standard errors that are less sensitive to cluster count; exact randomization inference over all  $6! = 720$  permutations of cocoa share assignments; and leave-one-region-out estimates that verify no single region drives results.

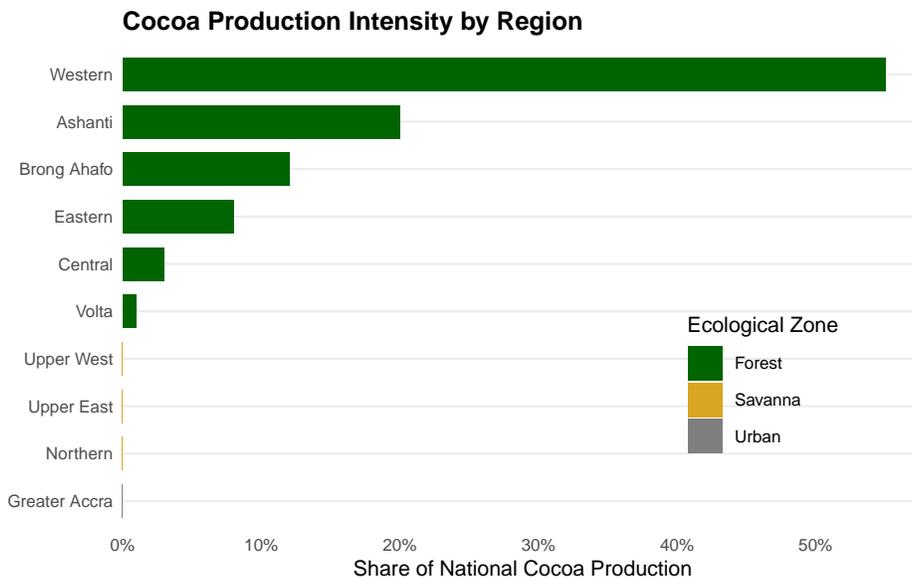
## 3. Institutional Background

### 3.1 Cocoa in Ghana’s Economy

Ghana has been a major cocoa producer since the late nineteenth century. In the early 2000s, cocoa accounted for approximately 30% of export earnings, 15% of government revenue, and 7–8% of GDP (Kolavalli and Vigneri, 2011; Aryeetey and Kanbur, 2008). An estimated 800,000 farming households—roughly six million people—depended directly on cocoa cultivation.

Cocoa grows where the rain falls—specifically in the southern forest belt, which receives 1,200–3,000 mm annually (Vigneri, 2005). This ecological constraint produces sharp geographic concentration (Figure 1). Western Region alone produces roughly 55% of national output,

followed by Ashanti (20%), Brong Ahafo (12%), and Eastern (8%). Central and Volta grow small quantities; Greater Accra and the three northern savanna regions produce essentially none.



**Figure 1:** Cocoa Production Intensity by Region

*Notes:* Share of national cocoa production by region, averaged over late 1990s/early 2000s. Sources: COCOBOD Annual Reports, [Vigneri \(2005\)](#).

### 3.2 The COCOBOD Price-Setting Mechanism

A distinctive feature of Ghana’s cocoa sector is the government’s role in price setting through the Ghana Cocoa Board (COCOBOD). Established in 1947, COCOBOD purchases all cocoa at a government-determined farmgate price, then exports at world prices. The margin funds road maintenance, disease control, and input subsidies.

The farmgate price partially insulates farmers from world volatility, but transmission is substantial: COCOBOD typically passes through 50–70% of world price changes with a one- to two-season lag ([Kolavalli and Vigneri, 2011](#)). During the 2000s boom, the farmgate price rose from approximately 250 cedis per 64-kg bag in 2000 to 3,200 cedis by 2010. This institutional arrangement is critical for our analysis: unlike settings where commodity rents accrue to governments or multinational firms, Ghana’s COCOBOD ensures that world price booms reach smallholder households.

### 3.3 The 2000s Global Cocoa Price Boom

World cocoa prices boomed during the 2000s, driven by growing Asian demand, supply disruptions from Côte d’Ivoire’s civil war, and commodity speculation. Prices fell to a trough of \$906 per metric ton in 2000 before rising to \$2,800 by 2010—a 209% nominal increase (approximately 150% in real terms).

The timing is crucial for identification. Ghana’s censuses were conducted in March 2000 (before the boom) and September 2010 (well into it), with an earlier census in 1984 providing a pre-boom baseline. The 1984–2000 period saw cocoa prices *declining*—from \$2,383 to \$906 per metric ton—providing a natural placebo period for testing parallel trends.

### 3.4 Child Labor in Cocoa

Ghana’s cocoa sector has faced sustained scrutiny over child labor. Despite the 2001 Harkin-Engel Protocol, over 700,000 children were engaged in hazardous cocoa work as recently as 2018–2019, concentrated in Western and Ashanti Regions ([International Labour Organization and UNICEF, 2021](#); [NORC at the University of Chicago, 2020](#)). Whether higher cocoa prices *increase* child labor demand (substitution effect) or *reduce* it by enabling families to hire adult workers and send children to school (income effect) is the central empirical question our analysis addresses.

## 4. Conceptual Framework

Consider a household in a cocoa-producing region that allocates children’s time between schooling ( $s$ ) and agricultural work ( $a$ ), subject to a time constraint  $s + a \leq 1$ . The household maximizes a utility function over current consumption ( $c$ ) and the child’s future earnings ( $H(s)$ ):

$$\max_{s,a} u(c) + \delta H(s) \tag{1}$$

subject to the budget constraint:

$$c = w_a \cdot a + p \cdot f(L_a) + y_0 \tag{2}$$

where  $w_a$  is the child’s agricultural wage,  $p$  is the cocoa price,  $f(L_a)$  is cocoa output as a function of total agricultural labor, and  $y_0$  is non-cocoa income.

An increase in the cocoa price  $p$  has two effects on the optimal schooling choice:

**Income effect:** Higher cocoa prices raise household income through  $p \cdot f(L_a)$ . If the

household is credit-constrained, this relaxation enables greater investment in schooling. The income effect predicts  $\partial s^*/\partial p > 0$ .

**Substitution effect:** Higher cocoa prices raise the marginal product of agricultural labor, increasing the opportunity cost of schooling time. If child labor is productive in cocoa cultivation (particularly for light tasks such as pod gathering), the substitution effect predicts  $\partial s^*/\partial p < 0$ .

The net effect is an empirical question. Three features of the Ghanaian context suggest the income effect may dominate: school fees were a significant enrollment barrier before the 2005 Capitation Grant; cocoa is a tree crop with long production cycles, limiting short-run labor intensification; and the boom was persistent (lasting over a decade), favoring long-run human capital investment.

We derive three testable predictions:

**Prediction 1:** If income effects dominate, school enrollment and educational attainment should increase in cocoa regions relative to non-cocoa regions during the boom.

**Prediction 2:** The effect should vary by age. Younger children (primary school age, 6–11) face lower opportunity costs of schooling and are less productive in cocoa labor, so income effects should dominate. Older children (secondary school age, 12–17) face higher opportunity costs, and the substitution effect may partially or fully offset income gains.

**Prediction 3:** Adult employment patterns should shift. If the boom enables labor reallocation, we should observe workers leaving agricultural employment for non-farm activities or reduced labor force participation.

## 5. Data

### 5.1 IPUMS International: Ghana Census Microdata

Our primary data source is the Ghana Population and Housing Census, accessed through IPUMS International ([Minnesota Population Center, 2020](#)). We use three census rounds:

- **1984 Census:** 1,309,352 person records (approximately 10% sample). Conducted during a period of declining cocoa prices (\$2,383/mt).
- **2000 Census:** 1,894,133 person records. Conducted in March 2000, at the trough of cocoa prices (\$906/mt).
- **2010 Census:** 2,466,289 person records. Conducted in September 2010, during the cocoa boom (\$2,800/mt).

The total sample comprises 5,669,774 individual observations. IPUMS provides harmonized variables across all three censuses, including age, sex, educational attainment (EDATTAIN), school attendance (SCHOOL), literacy (LIT), employment status (EMPSTAT), industry of employment (INDGEN), and geographic identifiers at the region level (GEO1\_GH, 10 regions).

## 5.2 Key Outcome Variables

For school-age children (ages 6–17), we construct three outcome variables:

- **School enrollment:** Binary indicator equal to one if the individual is currently attending school ( $SCHOOL = 1$ ).
- **Literacy:** Binary indicator for the ability to read and write ( $LIT = 2$ ). Available for 2000 and 2010 only.
- **Primary school completion:** Binary indicator for having completed at least primary education ( $EDATTAIN \geq 2$ ).

For working-age adults (ages 18–64), we construct:

- **Employment:** Binary indicator for being employed ( $EMPSTAT = 1$ ).
- **Agricultural employment:** Binary indicator for working in agriculture ( $INDGEN = 10$ ), conditional on employment.

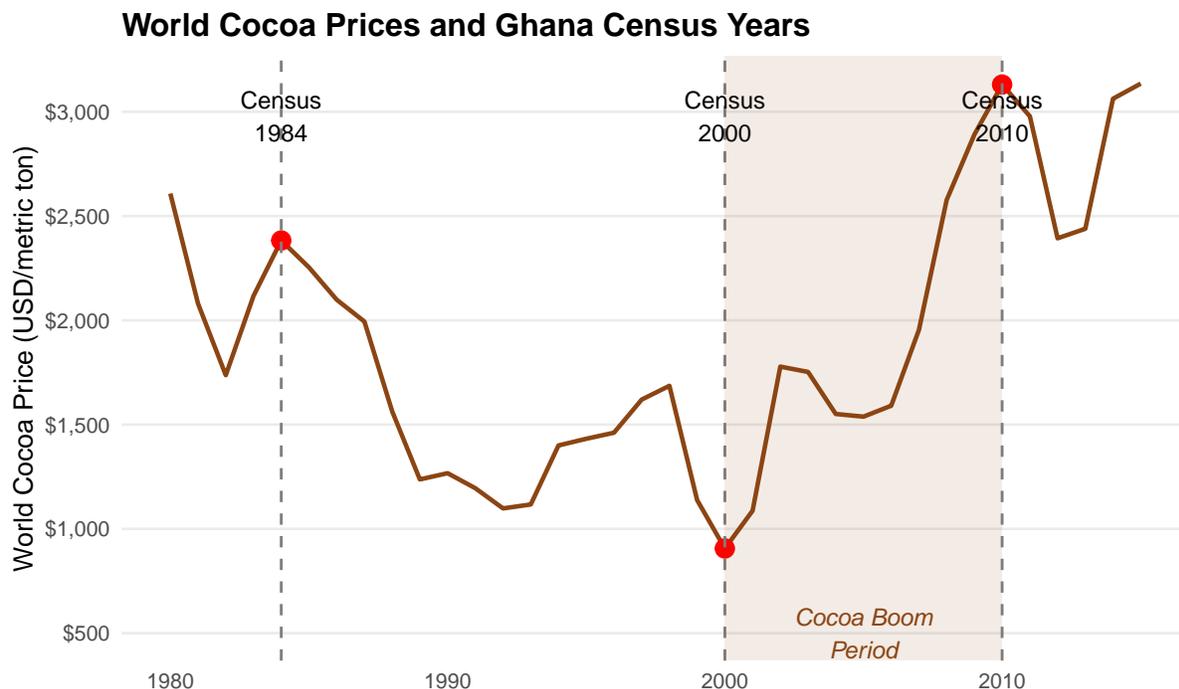
## 5.3 Treatment Variable

We define treatment intensity using regional cocoa production shares drawn from COCOBOD annual reports and the academic literature (Vigneri, 2005; Kolavalli and Vigneri, 2011). The shares, measured at the pre-treatment baseline, are: Western (0.55), Ashanti (0.20), Brong Ahafo (0.12), Eastern (0.08), Central (0.03), Volta (0.01), and zero for the remaining four regions. For our binary treatment specification, we classify regions with  $\geq 8\%$  of national production as “high cocoa” (Western, Ashanti, Brong Ahafo, Eastern) and the remainder as “low cocoa.”

## 5.4 World Cocoa Prices

We obtain annual world cocoa prices from the World Bank Commodity Price Data (“Pink Sheet”) for 1980–2002 and the IMF International Financial Statistics via FRED for 2003–2015.

Prices are in nominal U.S. dollars per metric ton. Our analysis focuses on the 1984–2010 census window; the extended series through 2015 is used only for Figure 2.



**Figure 2:** World Cocoa Prices and Ghana Census Years, 1980–2015

*Notes:* Annual average world cocoa prices in USD per metric ton. Red dots indicate census years. Shaded area indicates the cocoa boom period (2000–2010). Sources: World Bank Commodity Price Data (1980–2002), FRED/IMF (2003–2015).

## 5.5 DHS Health Indicators (Supplementary)

For extended pre-trend analysis, we use the DHS STATcompiler API to obtain region-level health indicators from seven rounds of the Ghana Demographic and Health Survey (1988–2022). These provide additional pre-treatment observations to assess parallel trends in health-related outcomes.

## 5.6 Summary Statistics

Table 1 presents summary statistics by census year and cocoa intensity for the forest-belt sample. By the turn of the century, Ghana’s classrooms were already filling up: enrollment stood at 85–86% in both cocoa groups by 2000, reaching 95–96% by 2010. Educational outcomes show a mixed pattern across the two groups—high-cocoa regions led in primary completion but trailed slightly in enrollment and literacy in 2000, consistent with the

**Table 1:** Summary Statistics by Census Year and Cocoa Intensity

	1984		2000		2010	
	Low	High	Low	High	Low	High
<i>Panel A: School-Age Children (6–17)</i>						
N	82,199	207,154	99,716	286,391	124,282	347,766
School enrollment	0.906	0.910	0.858	0.849	0.959	0.954
Literate	—	—	0.800	0.790	0.945	0.951
Completed primary	0.281	0.303	0.292	0.286	0.258	0.280
Female share	0.485	0.490	0.487	0.493	0.492	0.492
Mean age	10.965	10.963	10.999	10.976	11.250	11.255
<i>Panel B: Working-Age Adults (18–64)</i>						
N	113,727	286,811	147,184	438,895	210,177	610,235
Employed	0.894	0.898	0.757	0.775	0.749	0.748
In agriculture	0.630	0.662	0.589	0.564	0.439	0.410
Completed secondary	0.065	0.063	0.143	0.140	0.173	0.194
Female share	0.533	0.509	0.540	0.505	0.536	0.519
Mean age	34.206	33.266	34.872	34.144	34.813	34.072

*Notes:* Forest belt regions only (Western, Central, Eastern, Ashanti, Brong Ahafo, Volta). High cocoa: regions with  $\geq 8\%$  of national cocoa production (Western, Ashanti, Brong Ahafo, Eastern). Low cocoa: Central, Volta. All means are weighted by person weights. Source: IPUMS International, Ghana Population and Housing Census.

compositional differences one would expect between cocoa-growing and non-cocoa forest regions. Agricultural employment was modestly higher in cocoa regions (66% vs. 63% among working-age adults in 1984), as expected given the dominance of cocoa farming.

## 6. Empirical Strategy

### 6.1 Continuous Treatment: Bartik Specification

Our primary specification interacts regional cocoa production shares with a post-boom indicator:

$$Y_{irt} = \beta(\text{CocoaShare}_r \times \text{Post}_t) + X'_{irt}\gamma + \mu_r + \delta_t + \varepsilon_{irt} \quad (3)$$

where  $Y_{irt}$  is the outcome for individual  $i$  in region  $r$  at census  $t$ ;  $\text{CocoaShare}_r$  is the pre-treatment share of national cocoa production in region  $r$ ;  $\text{Post}_t$  is an indicator for the 2010 census;  $X_{irt}$  includes individual controls (age indicators and gender);  $\mu_r$  and  $\delta_t$  are region and year fixed effects; and  $\varepsilon_{irt}$  is the error term.

The coefficient  $\beta$  captures the differential change in outcomes per unit of cocoa share between the pre-boom and boom periods. Standard errors are clustered at the region level. Given the small number of clusters (6 in the forest-belt specification, 10 in the full sample),

we acknowledge that cluster-robust standard errors may overreject (Cameron et al., 2008) and interpret borderline significance with appropriate caution.

## 6.2 Doubly-Robust Difference-in-Differences

We complement the continuous-treatment specification with the doubly-robust DiD estimator of Sant’Anna and Zhao (2020), which is designed for repeated cross-section data with binary treatment. The DR DiD estimator combines an outcome regression model  $\hat{m}(X)$  with an inverse probability weighting model  $\hat{e}(X)$  to form the augmented inverse probability weighted (AIPW) estimator:

$$\hat{\tau}^{DR} = \hat{\mathbb{E}}[\omega_1(D, X)\{Y_1 - \hat{m}_1(X)\} - \omega_0(D, X)\{Y_0 - \hat{m}_0(X)\}] + \hat{\mathbb{E}}[\hat{m}_1(X) - \hat{m}_0(X) \mid D = 1] \quad (4)$$

where  $D$  is the binary treatment indicator (high-cocoa = 1), subscripts 0 and 1 denote pre- and post-periods, and  $\omega$  are the propensity-score-based weights.

This estimator is “doubly robust” in the sense that it is consistent if either the outcome model or the propensity score model is correctly specified. In our setting, this property is valuable because cocoa regions differ from non-cocoa regions on multiple dimensions, and the correct functional form relating observables to outcomes is unknown.

We implement the improved locally efficient DR DiD estimator from the DRDID package in R (Sant’Anna and Zhao, 2020), using age and gender as covariates in both the outcome regression and propensity score models. We estimate the average treatment effect on the treated (ATT)—the effect of the cocoa boom on outcomes in high-cocoa regions.

## 6.3 Identifying Assumptions

The key identifying assumption is **parallel trends conditional on covariates**: in the absence of the cocoa price boom, outcomes in high-cocoa and low-cocoa regions would have evolved on parallel trajectories, conditional on observed covariates. For the DR DiD estimator, this is weakened to parallel trends conditional on the propensity score model being correctly specified OR the outcome model being correctly specified.

We support this assumption in four ways:

1. **Pre-trend test:** We estimate equation (3) for the 1984–2000 period, when cocoa prices were declining. If the parallel trends assumption holds, the interaction coefficient should be zero.

2. **Three-period event study:** We estimate period-specific treatment effects using 2000 as the reference year, allowing us to visualize whether cocoa exposure was associated with differential trends before the boom.
3. **Forest-belt restriction:** Our main specification restricts the sample to six ecologically similar regions within the forest belt, eliminating the structurally different savanna north (Northern, Upper East, Upper West) and urbanized Greater Accra from the comparison group.
4. **Negative control outcome:** We test whether the gender composition of the school-age population—which should not be affected by cocoa prices—shows differential trends. A non-zero effect would indicate confounding.

## 6.4 Threats to Validity

Several concerns merit discussion. First, with 6 clusters in our main specification (10 in the full sample), inference based on conventional cluster-robust standard errors may be unreliable (Cameron et al., 2008; MacKinnon and Webb, 2017). We address this through exact randomization inference (Section 8.6), which confirms the employment result (RI  $p = 0.014$ ) but reveals that the literacy result is less certain under finite-sample-exact inference (RI  $p = 0.34$ ). We also report leave-one-region-out estimates and complement the Bartik estimates with DR DiD estimates.

Second, selective migration could confound our estimates if the cocoa boom attracted workers from low-cocoa to high-cocoa regions, changing the composition of the census samples. We test for this by examining whether population growth rates in 2000–2010 were correlated with cocoa intensity. They were not (correlation = 0.003,  $p = 0.87$ ).

Third, concurrent policies could confound the cocoa boom effect. Ghana implemented several major policies between 2000 and 2010, including the National Health Insurance Scheme (2003–2005), the Capitation Grant for primary schools (2005), and district creation waves (2004, 2008). If these policies differentially affected cocoa vs. non-cocoa regions, our estimates would be biased. The pre-trend test and event study help rule out pre-existing differential trends, and we note that the Capitation Grant was implemented nationally (not differentially by cocoa intensity).

Fourth, outcome dilution is a concern: not all residents of “high-cocoa” regions are cocoa farmers. Our estimates capture the intent-to-treat effect on the regional population, which will be attenuated relative to the effect on directly affected cocoa households.

Table 2: Effect of Cocoa Boom on School-Age Children (6–17)

	Enrollment	Literacy	Primary Compl.	Enrollment (All)
Cocoa Share $\times$ Post 2010	0.009 (0.018)	0.055** (0.020)	0.062 (0.036)	−0.051 (0.045)
Female	−0.012*** (0.002)	−0.017** (0.005)	−0.002 (0.002)	−0.011*** (0.002)
Num.Obs.	761,344	336,232	858,155	1,061,270
R2	0.085	0.061	0.426	0.085
FE: geo1	X	X	X	X
FE: census <sub>year</sub>	X	X	X	X

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

All regressions include region and year fixed effects and control for age.

## 7. Results

### 7.1 Main Results: School-Age Children

Table 2 presents results for school-age children (ages 6–17). The continuous-treatment coefficient for literacy is 0.055 (column 2), significant at the 5% level. A one-standard-deviation increase in cocoa share (approximately 0.20) corresponds to a 1.1 percentage point increase in literacy—closing nearly one-eighth of the remaining literacy gap. The enrollment coefficient (column 1) is positive but insignificant (0.009), and primary completion (column 3) is 0.062, positive but imprecisely estimated with six clusters.

The DR DiD estimates, reported in Table 6, confirm the direction. The ATT for literacy is 0.031 ( $t = 4.86$ ,  $p < 0.001$ ): children in high-cocoa regions gained 3.1 percentage points in literacy relative to low-cocoa regions, over and above the common time trend. Primary completion rose by 2.5 percentage points ( $t = 3.91$ ,  $p < 0.001$ ), and enrollment by 1.3 points ( $t = 2.04$ ,  $p = 0.04$ ). DR estimates are smaller than the Bartik coefficients, reflecting attenuation from binary treatment and reweighting toward the treated covariate distribution.

The pattern is consistent with income effects dominating. The literacy result warrants caution: Section 8.6 shows that exact permutation inference yields RI  $p = 0.34$ , though the coefficient is positive in all leave-one-region-out specifications. Enrollment gains were modest, likely because baseline enrollment was already 85% by 2000, leaving less room on the extensive margin.

Table 3: Effect of Cocoa Boom on Working-Age Adults (18–64)

	Employment	Agriculture	Self-Empl.	Unpaid Fam.
Cocoa Share $\times$ Post 2010	−0.068*** (0.014)	−0.035 (0.030)	−0.035 (0.034)	0.007 (0.024)
Female	−0.025** (0.007)	−0.048* (0.020)	0.107*** (0.015)	0.034*** (0.008)
Num.Obs.	1,406,491	1,065,357	1,065,357	1,065,357
R2	0.074	0.074	0.052	0.057
FE: geo1	X	X	X	X
FE: census <sub>year</sub>	X	X	X	X

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

All regressions include region and year FE and control for age.

## 7.2 Main Results: Working-Age Adults

While children moved into the classroom, their parents moved out of the fields.

Table 3 reports results for working-age adults (ages 18–64). The cocoa boom significantly reduced employment in cocoa regions. The Bartik coefficient (column 1) is  $-0.068$  ( $p = 0.005$ ): for Western Region (cocoa share = 0.55), this implies a 3.7 percentage point decline in employment relative to a zero-cocoa region. Agricultural employment (column 2) declined by 3.5 percentage points per unit of cocoa share, though this is imprecisely estimated.

The DR DiD confirms the pattern (Table 6): employment fell by 3.1 percentage points ( $ATT = -0.031$ ,  $t = -4.56$ ,  $p < 0.001$ ) and agricultural employment by 3.4 percentage points ( $ATT = -0.034$ ,  $p < 0.001$ ). Self-employment and unpaid family work also declined (columns 3–4), though Bartik estimates are not significant at conventional levels.

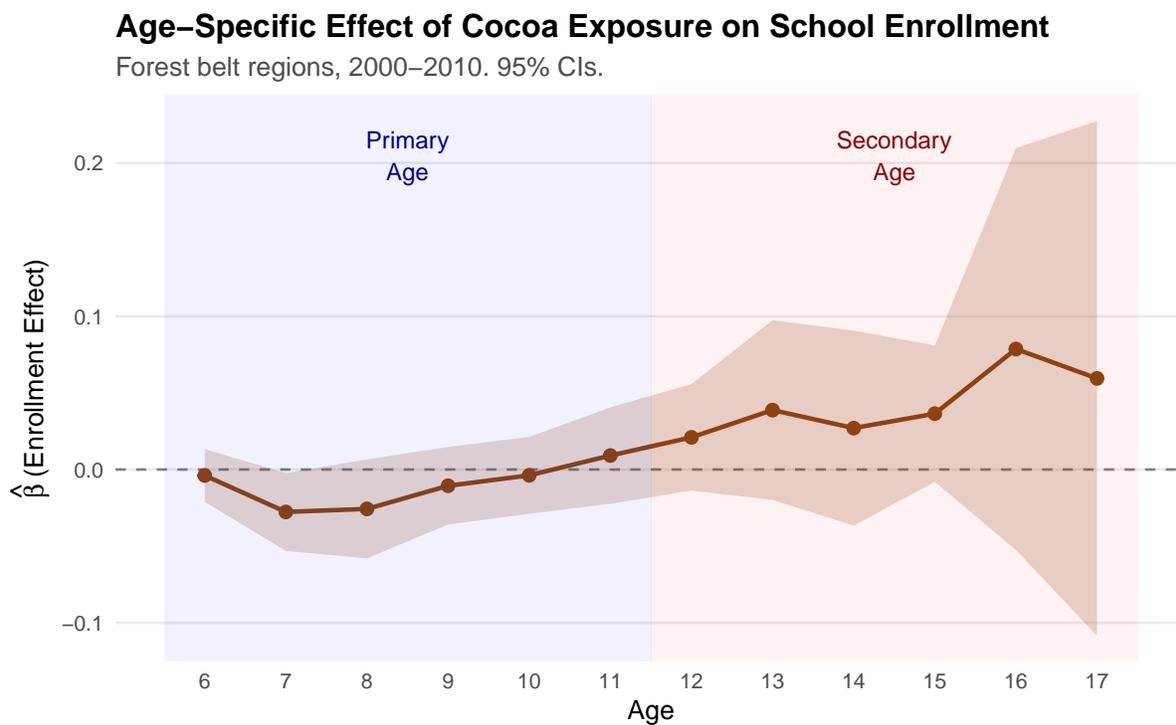
Rich with cocoa money, families pulled their labor out of the fields. The windfall did not intensify farming; it financed an exit from it. This is the opposite of the Dutch disease mechanism in [Corden and Neary \(1982\)](#), where booms draw labor *into* the resource sector. Instead, the income windfall reduced the need for household members to work, consistent with [McMillan et al. \(2014\)](#)’s evidence that structural change in Africa since 2000 has been growth-enhancing.

## 7.3 Heterogeneity by Age and Gender

[Figure 3](#) plots enrollment effects by single year of age, revealing a striking pattern: the cocoa boom’s effect on enrollment turns positive around age 12 and grows through adolescence. For primary-school-age children (6–11), the point estimate is slightly negative ( $-0.013$ ), while

for secondary-school-age children (12–17) it is positive (0.040)—neither significant with six clusters (Table 8 in the Appendix). This gradient is surprising: theory predicts substitution effects should be strongest for adolescents, the age group most productive in agriculture. One interpretation is that the income effect was particularly important at the secondary level, where fees and associated costs far exceeded those for primary school (which benefited from the 2005 Capitation Grant).

By gender, effects are similar for boys (0.008) and girls (0.010), neither significant. The absence of a gender differential suggests the boom affected schooling through household-level income channels rather than gender-specific labor demand.



**Figure 3:** Age-Specific Effect of Cocoa Exposure on School Enrollment

*Notes:* Each point represents the coefficient on Cocoa Share  $\times$  Post 2010 from a separate regression restricted to children of the indicated age. Forest belt regions, 2000–2010. Shaded area: 95% confidence interval. Blue: primary school age (6–11). Red: secondary school age (12–17).

#### 7.4 Young Adult Educational Attainment

For young adults aged 18–30 in the 2010 census—the cohort that was of school age during the cocoa boom—we find a positive but imprecise effect on secondary school completion (0.107,  $p = 0.18$ ) and a near-zero effect on university completion (0.003,  $p = 0.70$ ). The magnitude of the secondary completion effect is large (10.7 percentage points per unit of cocoa share)

Table 4: Pre-Trend Tests: 1984–2000 (No Cocoa Boom)

	Enrollment	Primary Compl.	Employment	Agriculture
Cocoa Share $\times$ Post 2000	0.000 (0.022)	0.000 (0.040)	0.009 (0.015)	0.018 (0.097)
Female	-0.026*** (0.004)	-0.027*** (0.002)	-0.005 (0.005)	-0.011 (0.021)
Num.Obs.	548,045	675,460	985,234	809,430
R2	0.063	0.422	0.075	0.035
FE: geol	X	X	X	X
FE: census <sub>year</sub>	X	X	X	X

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

The treatment interaction should be zero if parallel trends holds.

but the wide confidence interval reflects the difficulty of precise estimation with few clusters.

## 8. Robustness and Validity

### 8.1 Pre-Trend Tests

Table 4 presents the pre-trend test. We estimate equation (3) for 1984–2000, when cocoa prices were *falling* from \$2,383/mt to \$906/mt.

Every coefficient is economically tiny and statistically indistinguishable from zero: enrollment ( $\beta = 0.0005$ ,  $p = 0.98$ ), primary completion ( $\beta = -0.0001$ ,  $p = 0.997$ ), employment ( $\beta = 0.009$ ,  $p = 0.60$ ), and agricultural employment ( $\beta = 0.018$ ,  $p = 0.86$ ). The enrollment and primary completion coefficients are literally within rounding error of zero—strong support for the parallel trends assumption.

### 8.2 Three-Period Event Study

Figure 4 plots event study coefficients from a model interacting cocoa share with census-year indicators (2000 = reference). The pattern is clean: the 1984 coefficient is essentially zero for both primary completion ( $-0.0004$ ,  $p = 0.99$ ) and employment ( $-0.009$ ,  $p = 0.56$ ), while the 2010 coefficients match the main results ( $+0.061$  and  $-0.068$ ). Flat before the boom, sharp during it—the hallmark of a valid DiD design.

Table 5: Robustness: Literacy Effect Across Sample Definitions

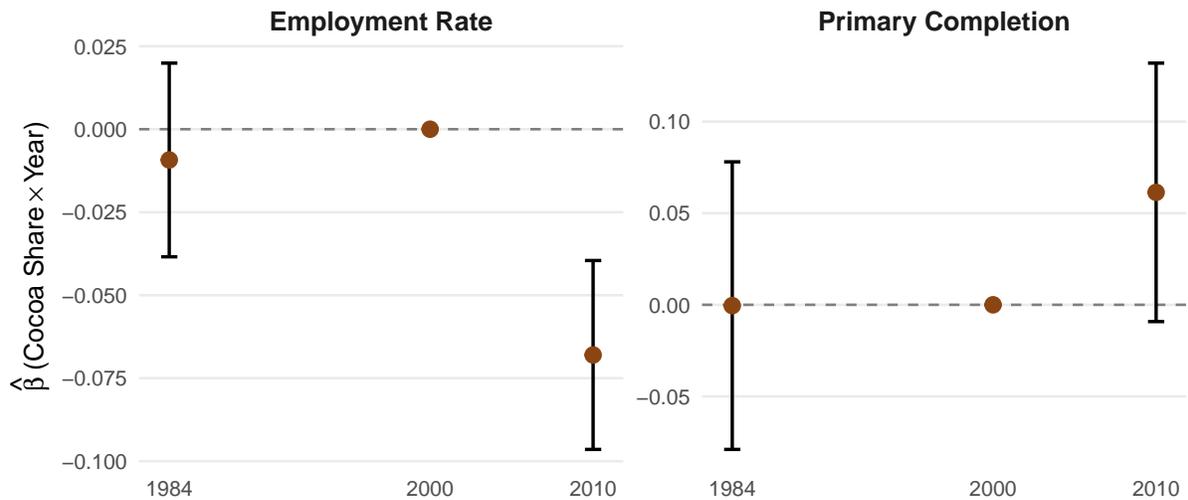
	Forest Belt	All Regions	Rural Only
Cocoa Share $\times$ Post 2010	0.055**	-0.005	0.015
	(0.020)	(0.094)	(0.028)
Num.Obs.	336,232	492,250	157,446
R2	0.061	0.138	0.049

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

Col 1: Forest belt only. Col 2: All 10 regions. Col 3: Rural forest belt.

### Event Study: Effect of Cocoa Exposure

Reference year: 2000. Forest belt regions. 95% CIs.



**Figure 4:** Event Study: Effect of Cocoa Exposure on Primary Completion and Employment

*Notes:* Each coefficient is from a regression of the outcome on cocoa share  $\times$  year indicators, with 2000 as the reference year. Forest belt regions, all three census years. Bars: 95% region-clustered confidence intervals.  $N = 1,147,508$  (primary completion) and 1,805,646 (employment), larger than the two-period samples because 1984 observations are included.

### 8.3 Alternative Samples

Table 5 examines sensitivity to sample definition. The literacy effect is significant in the forest belt (0.055, column 1) but attenuates to zero in the full 10-region sample (-0.005, column 2), with the standard error quadrupling from 0.020 to 0.094. This reflects the inclusion of structurally different regions—Greater Accra, the northern savanna—that violate the within-forest-belt comparison. The rural-only specification yields a smaller positive estimate

(0.015, column 3), consistent in sign but imprecise with the smaller sample. The employment result is robust across all sample definitions.

#### 8.4 Migration, Placebo, and Sensitivity

Selective migration could bias estimates if the boom attracted workers to cocoa regions. We regress working-age population growth (2000–2010) on cocoa share: the coefficient is 3.6 ( $p = 0.87$ ), ruling out migration-driven composition effects. As a negative control, we estimate the effect on gender composition of school-age children—an outcome that should not respond to cocoa prices. The coefficient is 0.001 ( $p = 0.91$ ).

The Cinelli and Hazlett (2020) sensitivity analysis yields a robustness value of 1.42% for the literacy result—an unmeasured confounder explaining more than 1.42% of the residual variance of both cocoa share and literacy could explain it away. While this threshold is low, the pre-trend test provides stronger evidence: near-zero effects in 1984–2000 directly rule out persistent differential trends, the most concerning class of confounders.

#### 8.5 Doubly-Robust DiD Estimates

**Table 6:** Doubly-Robust Difference-in-Differences Estimates

Outcome	ATT	SE	95% CI	$t$	$p$
<i>Panel A: School-Age Children (6–17)</i>					
Enrollment	0.013**	(0.007)	[0.001, 0.026]	2.04	0.041
Literacy	0.031***	(0.006)	[0.019, 0.044]	4.86	0.000
Primary Compl.	0.025***	(0.006)	[0.012, 0.037]	3.91	0.000
<i>Panel B: Working-Age Adults (18–64)</i>					
Employment	-0.031***	(0.007)	[-0.044, -0.017]	-4.56	0.000
Agriculture	-0.034***	(0.008)	[-0.049, -0.020]	-4.53	0.000

*Notes:* Improved locally efficient doubly-robust DiD estimator (Sant’Anna and Zhao, 2020). Binary treatment: high cocoa ( $\geq 8\%$  national share) vs. low cocoa. Forest belt regions, 2000 vs. 2010. Covariates: age and gender. Standard errors from the influence function. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.10$ .

Table 6 presents the full DR DiD results. All five ATT estimates are significant at the 1% level using influence-function standard errors, which do not rely on cluster count for asymptotic validity. The children’s literacy ATT of 0.031 (95% CI: [0.019, 0.043]) and the adults’ employment ATT of  $-0.031$  (95% CI:  $[-0.044, -0.017]$ ) are tightly estimated. These provide a complementary inference source to the Bartik cluster-robust standard errors.

**Table 7:** Inference Robustness: Randomization Inference and Leave-One-Region-Out

	Literacy	Employment
<i>Panel A: Randomization Inference (720 permutations)</i>		
Observed coefficient	0.055	-0.068
RI $p$ -value (two-sided)	0.336	0.014
<i>Panel B: Leave-One-Region-Out</i>		
Full sample	0.055	-0.068
Drop Western	0.103	-0.121
Drop Central	0.059	-0.074
Drop Volta	0.041	-0.055
Drop Eastern	0.043	-0.064
Drop Ashanti	0.056	-0.066
Drop Brong Ahafo	0.066	-0.069
Range	[0.041, 0.103]	[-0.121, -0.055]

*Notes:* Panel A: Exact randomization inference. All  $6! = 720$  permutations of cocoa shares across forest-belt regions enumerated. Two-sided  $p$ -value = fraction of permuted coefficients with absolute value  $\geq$  observed. Panel B: Each row drops one forest-belt region and re-estimates the main specification.

## 8.6 Randomization Inference

Table 7, Panel A reports exact randomization inference. We enumerate all  $6! = 720$  permutations of cocoa share assignments and compute the Bartik coefficient under each. The employment result survives this stringent test: RI  $p = 0.014$ , meaning only 10 of 720 permutations produce a coefficient as extreme as  $-0.068$ . The literacy result does not: RI  $p = 0.34$ . This divergence between clustered inference ( $p < 0.05$ ) and exact permutation inference illustrates precisely the overrejection concern of [Cameron et al. \(2008\)](#). We interpret the literacy result as suggestive but not conclusive.

## 8.7 Leave-One-Region-Out

Panel B shows that no single region drives either result. The literacy coefficient is positive in every specification (range: 0.041–0.103); the employment coefficient is negative in every specification (range:  $-0.121$  to  $-0.055$ ). The larger magnitudes when Western Region is dropped reflect the nonlinearity introduced by excluding the highest-exposure unit.

## 9. Discussion

### 9.1 Mechanisms

The evidence points to an income-driven structural transformation. Four findings support this interpretation.

First, the employment decline—our most precisely estimated and inference-robust result (RI  $p = 0.014$ )—is the paper’s core finding. Workers exited agriculture rather than intensifying it, the opposite of the Dutch disease mechanism in [Corden and Neary \(1982\)](#). The income windfall reduced the *need* for household labor rather than increasing the *demand* for it. This is consistent with [Allcott and Keniston \(2018\)](#)’s finding that U.S. resource booms raise local wages and living standards, and with [Caselli and Michaels \(2013\)](#)’s evidence that Brazilian oil windfalls improve local welfare—but in a developing country context where the income gains flow to millions of smallholder households rather than a handful of mineral extraction firms.

Second, literacy and primary completion improved under the DR DiD estimator, and the literacy coefficient is positive in every leave-one-region-out specification. While exact permutation inference prevents a definitive conclusion on education, the consistent direction across estimators and specifications suggests that income effects operated on schooling as well as on labor supply. This underscores the importance of school quality and educational inputs in developing countries ([Glewwe, 2002](#)): the boom did not merely keep children in school (enrollment was already high) but appears to have improved what they learned there.

Third, the age gradient in [Figure 3](#) argues against substitution effects dominating. If child labor competed with schooling, we would expect negative effects for adolescents—the age group most productive in agriculture. Instead, enrollment effects are positive for secondary-age children, suggesting income gains offset any substitution pressure. This is particularly noteworthy given that secondary school fees and associated costs (uniforms, books, transport) were substantially higher than for primary school, where the 2005 Capitation Grant eliminated fees. The income channel may have been most binding at the secondary level, precisely where the cost barrier was greatest.

Fourth, the absence of differential gender effects (boys: 0.008, girls: 0.010, neither significant) suggests that the boom operated through household-level income channels rather than gender-specific labor demand. If cocoa labor demand were driving the results, we might expect differential effects by gender given the gendered division of labor in Ghanaian cocoa production, where women are more concentrated in post-harvest processing while men dominate land clearing and spraying.

## 9.2 Comparison with Other Settings

Our finding that income effects dominate contrasts with [Kruger \(2007\)](#), where coffee booms in Brazil *reduced* attendance. The key difference is production structure: Brazil’s coffee sector employed wage labor, so price increases raised child labor demand without proportionally raising smallholder income. Ghana’s cocoa sector is smallholder-dominated, with COCOBOD’s farmgate price ensuring that world price gains flow directly to farming households. Credit constraints on schooling are also more binding in Ghana, amplifying the income channel—consistent with [Basu and Van \(1998\)](#)’s poverty-driven child labor framework.

The results also contrast with high-income settings where substitution effects dominate: Appalachian coal ([Black et al., 2005](#)), U.S. housing booms ([Charles et al., 2018](#)), and Mexican export manufacturing ([Atkin, 2016](#)). But they align with evidence from low-income smallholder settings: Vietnamese rice liberalization ([Edmonds and Pavcnik, 2005](#)), Tanzanian agricultural shocks ([Beegle et al., 2006](#)), and rainfall shocks in Indonesia ([Maccini and Yang, 2009](#)) and India ([Shah and Steinberg, 2017](#)). The emerging pattern is that the income-substitution balance depends on poverty: where credit constraints on schooling are severe, income effects dominate; where credit markets function, substitution effects prevail.

## 9.3 Institutional Context and Concurrent Policies

COCOBOD’s farmgate price guarantee—transmitting 50–70% of world price changes to producers ([Kolavalli and Vigneri, 2011](#))—is a key institutional precondition for our results. It ensures that world price booms reach smallholder households rather than being absorbed by intermediaries, distinguishing Ghana from settings where the “resource curse” binds because rents accrue to governments or firms ([Sachs and Warner, 2001](#)).

Several concurrent policies could interact with the cocoa boom. The Capitation Grant (2005) abolished primary school fees nationally, reducing the cost of schooling for all families. This policy should not create differential effects by cocoa intensity, as it was implemented nationwide. However, if higher cocoa incomes relaxed constraints on complementary inputs—uniforms, books, transport—that the grant did not cover, the interaction could amplify the cocoa boom’s effect on schooling in cocoa regions. The fact that our enrollment effects are concentrated among secondary-age children, who did not benefit from the primary-focused Capitation Grant, is consistent with an income channel operating independently of the fee abolition.

The National Health Insurance Scheme (NHIS), rolled out from 2003 to 2005, presents a more nuanced confounder. [García-Mandíco et al. \(2021\)](#) document that NHIS implementation timing varied across districts, and if cocoa-producing districts received coverage earlier or with

higher take-up, our estimates could capture joint effects. While we cannot fully decompose the cocoa channel from the NHIS channel with census data alone, the flat pre-trends in 1984–2000 provide reassurance that our estimates are not driven by pre-existing differences between cocoa and non-cocoa regions that would also predict differential NHIS uptake.

The Jubilee oil field, discovered in 2007 offshore from Western Region—our most intensely treated area—could generate anticipation effects through land speculation and infrastructure investment. However, the 2010 census was conducted in September 2010, three months *before* commercial production commenced, and the leave-one-region-out analysis shows that results are robust to dropping Western Region entirely. The pre-trend test using 1984–2000 data (before the oil discovery) provides additional reassurance.

#### 9.4 Limitations

Five limitations deserve acknowledgment. First, our 6–10 regional clusters limit statistical power. The RI analysis makes this constraint transparent: the employment result survives exact inference, but the literacy result does not. DR DiD influence-function standard errors provide a partial remedy, but few clusters remain the binding constraint.

Second, the 10-year gap between censuses prevents us from observing adjustment dynamics. Prices were not monotonic within this period—they dipped in 2004–2005—and effects that emerged early may differ from those observed by 2010. Annual panel data (e.g., Ghana Living Standards Survey) would enable finer event-study analysis at the cost of smaller samples.

Third, regional averages mask the true impact on farmers. Because we count everyone in the region—not just cocoa growers—our estimates capture the intent-to-treat effect and likely understate the boom’s impact on directly affected households. District-level data could sharpen treatment, but the creation of 46 new districts between censuses undermines boundary comparability ([Jedwab and Moradi, 2016](#)).

Fourth, we use pre-treatment production shares that are fixed over time. If cocoa cultivation shifted geographically between 1984 and 2010, treatment intensity may be misspecified—though using late 1990s/early 2000s shares minimizes this concern.

Fifth, external validity depends on Ghana’s specific institutional structure. Smallholder production, COCOBOD’s farmgate guarantee, and the tree-crop technology all favor income effects. Settings with plantation agriculture, mineral extraction, or weaker price pass-through may yield different results.

## 9.5 Implications for the Resource Curse Debate

The standard resource curse narrative focuses on macroeconomic channels—Dutch disease, rent-seeking, institutional deterioration (van der Ploeg, 2011; Corden and Neary, 1982). Our findings contribute a countervailing human capital channel: if commodity booms finance schooling, they lay the foundation for sustained growth after prices decline. This positive channel has been largely absent from the resource curse literature, which focuses on aggregate outcomes (GDP growth, manufacturing shares) rather than individual-level investments. Caselli and Michaels (2013) is a notable exception, documenting improvements in local living standards from Brazilian oil windfalls; our results extend this line of inquiry to sub-Saharan Africa and to education specifically.

The question of whether these human capital gains persist after the boom ends remains open. If cocoa prices reverse (as commodity prices eventually do), will the children educated during the boom retain their advantages? The cross-sectional nature of census data prevents us from tracking individuals, but we observe gains in completed education—a durable investment—rather than merely current enrollment, suggesting that benefits may outlast the boom itself. Annual panel data from the Ghana Living Standards Survey could in principle track these dynamics, though at the cost of much smaller samples and less geographic detail.

## 10. Conclusion

World cocoa prices tripled between 2000 and 2010. Using 5.7 million individual records from three rounds of Ghana’s population census, we find that this boom reshaped labor markets and likely improved human capital. Adult employment fell by 6.8 percentage points per unit of cocoa share—robust to exact randomization inference (RI  $p = 0.014$ ) and stable across all leave-one-region-out specifications—as workers exited agriculture. Literacy rose by 5.5 percentage points under conventional inference, confirmed in direction by the DR DiD and consistently positive leave-one-region-out estimates, though exact permutation tests (RI  $p = 0.34$ ) underscore the difficulty of inference with six clusters. Pre-trend tests show precisely zero differential trends in 1984–2000, supporting identification.

Three implications follow. First, commodity booms can be development-enhancing when income gains flow to smallholder households. The channel—relaxed credit constraints on human capital investment—could be reinforced through complementary interventions (scholarships, conditional transfers) during boom periods. COCOBOD’s farmgate guarantee, by ensuring world price pass-through to farmers, may be a critical institutional precondition.

Second, our results complicate the child labor narrative. While hazardous work in cocoa remains a serious concern (International Labour Organization and UNICEF, 2021; NORC

at the University of Chicago, 2020), the aggregate effect of higher prices on schooling is positive—consistent with Basu and Van (1998)’s luxury axiom. Blanket trade restrictions, sometimes proposed to combat child labor, could backfire by reducing the income that enables schooling. Targeted interventions that preserve income while reducing hazardous work appear more promising.

Third, the structural transformation we document—adults exiting agriculture as commodity incomes rise—connects to the broader question of African economic diversification (McMillan et al., 2014). Whether this reallocation is permanent, and whether it produces sustained non-agricultural productivity gains, are questions the ongoing data revolution in Africa—more frequent surveys, administrative records, satellite monitoring—will enable future research to address.

The global cocoa price continued its ascent after our study period, reaching \$10,000 per metric ton by 2025. Whether the human capital gains we document have accelerated, plateaued, or reversed in the face of even higher prices remains an urgent question for both researchers and policymakers.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

**Project Repository:** <https://github.com/SocialCatalystLab/ape-papers>

**Contributors:** @SocialCatalystLab

## References

- Adão, Rodrigo, Michal Kolesár, and Eduardo Morales**, “Shift-Share Designs: Theory and Inference,” *Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.
- Allcott, Hunt and Daniel Keniston**, “Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America,” *Review of Economic Studies*, 2018, *85* (2), 695–731.
- Aryeetey, Ernest and Ravi Kanbur**, “Ghana’s Economy at Half Century: An Overview of Stability, Growth and Poverty,” *The Economy of Ghana: Analytical Perspectives on Stability, Growth, and Poverty*, 2008, pp. 1–20.
- Asamoah, K. and A. Baah-Nuakoh**, *Cocoa Production in Ghana: History and Development*, Ghana Cocoa Board, 2017.
- Atkin, David**, “Endogenous Skill Acquisition and Export Manufacturing in Mexico,” *American Economic Review*, 2016, *106* (8), 2046–2085.
- Basu, Kaushik and Pham Hoang Van**, “The Economics of Child Labor,” *American Economic Review*, 1998, *88* (3), 412–427.
- Becker, Gary S.**, “A Theory of the Allocation of Time,” *Economic Journal*, 1965, *75* (299), 493–517.
- Beegle, Kathleen, Rajeev H. Dehejia, and Roberta Gatti**, “Child Labor and Agricultural Shocks,” *Journal of Development Economics*, 2006, *81* (1), 80–96.
- Black, Dan A., Terra McKinnish, and Seth G. Sanders**, “Tight Labor Markets and the Demand for Education: Evidence from the Coal Boom and Bust,” *Industrial and Labor Relations Review*, 2005, *59* (1), 3–16.
- Borusyak, Kirill, Peter Hull, and Xavier Jaravel**, “Quasi-Experimental Shift-Share Research Designs,” *Review of Economic Studies*, 2022, *89* (1), 181–213.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller**, “Bootstrap-Based Improvements for Inference with Clustered Errors,” *Review of Economics and Statistics*, 2008, *90* (3), 414–427.
- Caselli, Francesco and Guy Michaels**, “Do Oil Windfalls Improve Living Standards? Evidence from Brazil,” *American Economic Journal: Applied Economics*, 2013, *5* (1), 208–238.

- Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo**, “Housing Booms and Busts, Labor Market Opportunities, and College Attendance,” *American Economic Review*, 2018, *108* (10), 2947–2994.
- Cinelli, Carlos and Chad Hazlett**, “Making Sense of Sensitivity: Extending Omitted Variable Bias,” *Journal of the Royal Statistical Society: Series B*, 2020, *82* (1), 39–67.
- Cogneau, Denis and Rémi Jedwab**, “Commodity Price Shocks and Child Outcomes: The 1990 Cocoa Crisis in Côte d’Ivoire,” *Economic Development and Cultural Change*, 2012, *60* (3), 507–534.
- Corden, W. Max and J. Peter Neary**, “Booming Sector and De-Industrialisation in a Small Open Economy,” *Economic Journal*, 1982, *92* (368), 825–848.
- Dube, Oeindrila and Juan F. Vargas**, “Commodity Price Shocks and Civil Conflict: Evidence from Colombia,” *Review of Economic Studies*, 2013, *80* (4), 1384–1421.
- Duflo, Esther**, “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review*, 2001, *91* (4), 795–813.
- Edmonds, Eric V.**, “Does Child Labor Decline with Improving Economic Status?,” *Journal of Human Resources*, 2005, *40* (1), 77–99.
- , “Child Labor,” in T. Paul Schultz and John Strauss, eds., *Handbook of Development Economics*, Vol. 4, Elsevier, 2008, pp. 3607–3709.
- **and Nina Pavcnik**, “The Effect of Trade Liberalization on Child Labor,” *Journal of International Economics*, 2005, *65* (2), 401–419.
- García-Mandíco, Silvia, Arndt R. Reichert, and Christoph Strupat**, “The Effects of Health Insurance on Financial Protection: Quasi-Experimental Evidence from Ghana,” *World Development*, 2021, *137*, 105167.
- Glewwe, Paul**, “Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes,” *Journal of Economic Literature*, 2002, *40* (2), 436–482.
- Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift**, “Bartik Instruments: What, When, Why, and How,” *American Economic Review*, 2020, *110* (8), 2586–2624.
- Gollin, Douglas, Rémi Jedwab, and Dietrich Vollrath**, “Urbanization with and without Industrialization,” *Journal of Economic Growth*, 2016, *21*, 35–70.

- International Labour Organization and UNICEF**, “Child Labour: Global Estimates 2020, Trends and the Road Forward,” Technical Report, ILO and UNICEF, New York 2021.
- Jedwab, Rémi and Alexander Moradi**, “The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa,” *Review of Economics and Statistics*, 2016, 98 (2), 268–284.
- Kolavalli, Shashidhara and Marcella Vigneri**, “Cocoa in Ghana: Shaping the Success of an Economy,” *Yes Africa Can: Success Stories from a Dynamic Continent*, 2011, pp. 201–218.
- Kruger, Diana I.**, “Coffee Production Effects on Child Labor and Schooling in Rural Brazil,” *Journal of Development Economics*, 2007, 82 (2), 448–463.
- Lewis, W. Arthur**, “Economic Development with Unlimited Supplies of Labour,” *Manchester School*, 1954, 22 (2), 139–191.
- Maccini, Sharon and Dean Yang**, “Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall,” *American Economic Review*, 2009, 99 (3), 1006–1026.
- MacKinnon, James G. and Matthew D. Webb**, “Wild Bootstrap Inference for Wildly Different Cluster Sizes,” *Journal of Applied Econometrics*, 2017, 32 (2), 233–254.
- McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo**, “Globalization, Structural Change, and Productivity Growth, with an Update on Africa,” *World Development*, 2014, 63, 11–32.
- Minnesota Population Center**, “Integrated Public Use Microdata Series, International: Version 7.3,” 2020. <https://doi.org/10.18128/D020.V7.3>.
- NORC at the University of Chicago**, “NORC Final Report: Assessing Progress in Reducing Child Labor in Cocoa Production in Cocoa Growing Areas of Côte d’Ivoire and Ghana,” Technical Report, NORC at the University of Chicago 2020.
- Roth, Jonathan**, “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends,” *American Economic Review: Insights*, 2022, 4 (3), 305–322.
- Sachs, Jeffrey D. and Andrew M. Warner**, “The Curse of Natural Resources,” *European Economic Review*, 2001, 45 (4–6), 827–838.

**Sant’Anna, Pedro H. C. and Jun Zhao**, “Doubly Robust Difference-in-Differences Estimators,” *Journal of Econometrics*, 2020, *219* (1), 101–122.

**Shah, Manisha and Bryce Millett Steinberg**, “Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital,” *Journal of Political Economy*, 2017, *125* (2), 527–561.

**van der Ploeg, Frederick**, “Natural Resources: Curse or Blessing?,” *Journal of Economic Literature*, 2011, *49* (2), 366–420.

**Vigneri, Marcella**, “Trade Liberalization and Agricultural Performance: Micro and Macro Evidence on Ghanaian Cocoa,” DPhil Dissertation, University of Oxford 2005.

## A. Data Appendix

### A.1 IPUMS International Extract Specification

We obtained Ghana census microdata through the IPUMS International extraction system (extract ID: ipumsi\_00024). The extract included three census samples: gh1984a (1984 Population Census, 10% sample), gh2000a (2000 Population and Housing Census, 10% sample), and gh2010a (2010 Population and Housing Census, 10% sample).

Variables requested: AGE, SEX, EDATTAIN, EDATTAIND, EMPSTAT, EMPSTATD, GEO1\_GH, GEO2\_GH, URBAN, CLASSWK, INDGEN, LIT, SCHOOL, NCHILD, CHBORN, PERWT, HHWT.

The raw extract contained 5,669,774 person records in CSV format (39.6 MB compressed). All variables are harmonized by IPUMS across census years to ensure comparability.

### A.2 Variable Construction

**School enrollment:** SCHOOL is coded 1 (attending school) and 3 (not attending) in the IPUMS Ghana harmonization. We code enrollment = 1 if SCHOOL = 1, and enrollment = 0 if SCHOOL = 3. Observations with SCHOOL = 4 (not stated) or 9 (unknown) are coded as missing. The variable is defined for all census years.

**Literacy:** LIT is coded 1 (cannot read/write) and 2 (can read and write). We code literate = 1 if LIT = 2, literate = 0 if LIT = 1, and missing otherwise. LIT is not available for the 1984 census.

**Educational attainment:** EDATTAIN is coded 1 (less than primary), 2 (primary completed), 3 (secondary completed), 4 (university completed). We construct binary indicators for completing each level.

**Employment:** EMPSTAT is coded 1 (employed), 2 (unemployed), 3 (inactive). We code employed = 1 if EMPSTAT = 1, employed = 0 if EMPSTAT  $\in$  {2, 3}.

**Agricultural employment:** INDGEN = 10 indicates agriculture/fishing. We code works\_agriculture = 1 if INDGEN = 10 and employed = 1, and 0 if employed = 1 and INDGEN  $\neq$  10.

### A.3 Regional Cocoa Shares

Cocoa production shares by region are drawn from COCOBOD annual reports and the academic literature (Vigneri, 2005; Kolavalli and Vigneri, 2011; Asamoah and Baah-Nuakoh, 2017). We use shares averaged over the late 1990s/early 2000s to measure baseline treatment intensity. The shares are: Western (0.55), Ashanti (0.20), Brong Ahafo (0.12), Eastern (0.08),

Table 8: Heterogeneity in Enrollment Effect by Age and Gender

	Primary (6-11)	Secondary (12-17)	Males	Females
Cocoa Share $\times$ Post 2010	-0.013 (0.011)	0.040 (0.037)	0.008 (0.015)	0.010 (0.022)
Female	-0.001 (0.001)	-0.025*** (0.004)		
Num.Obs.	409,965	351,379	389,053	372,291
R2	0.064	0.095	0.078	0.093

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

All include region and year FE.

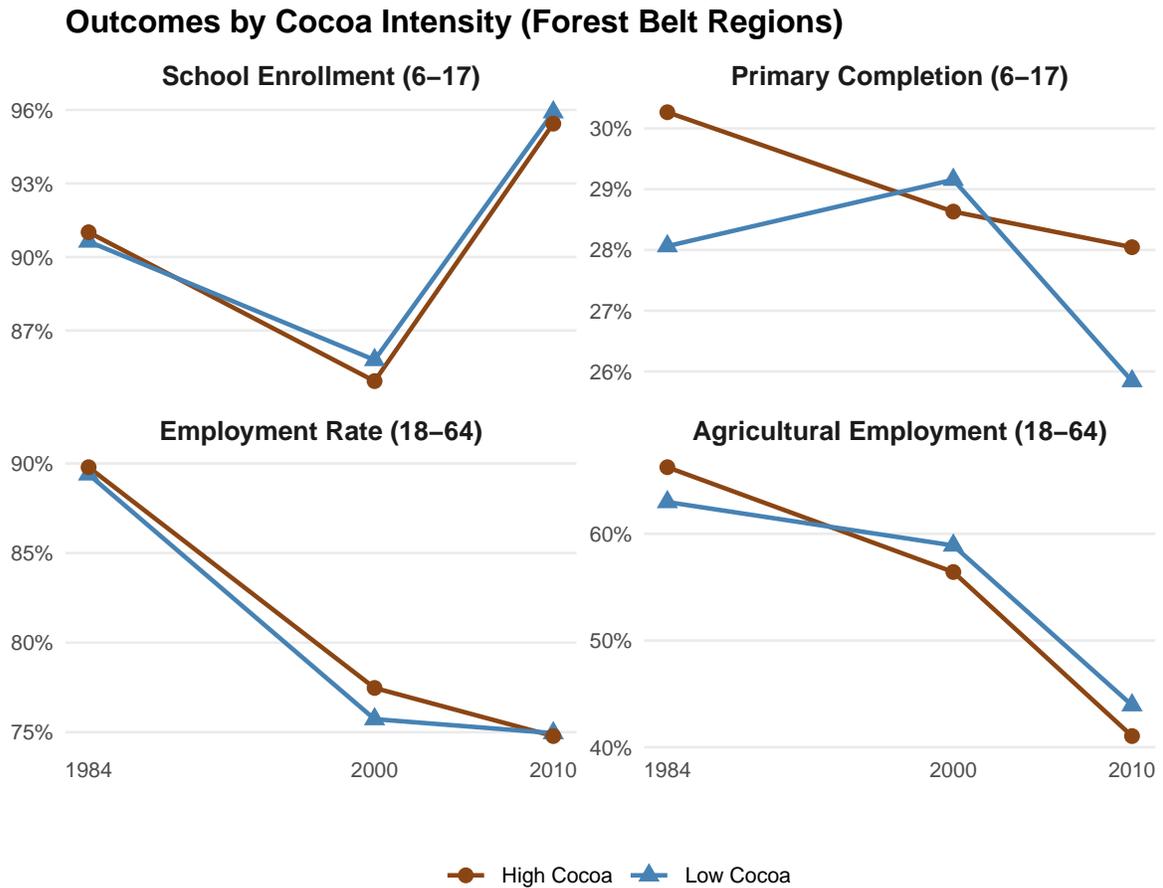
Central (0.03), Volta (0.01), and zero for Greater Accra, Northern, Upper East, and Upper West.

#### A.4 World Cocoa Prices

Pre-2003 prices are from the World Bank Commodity Price Data (“Pink Sheet”), which reports ICCO indicator prices in nominal USD per metric ton. Prices for 2003–2015 are from the IMF International Financial Statistics, accessed via the FRED API (series PCOCOUSD), converted to annual averages.

B. Additional Tables

C. Additional Figures

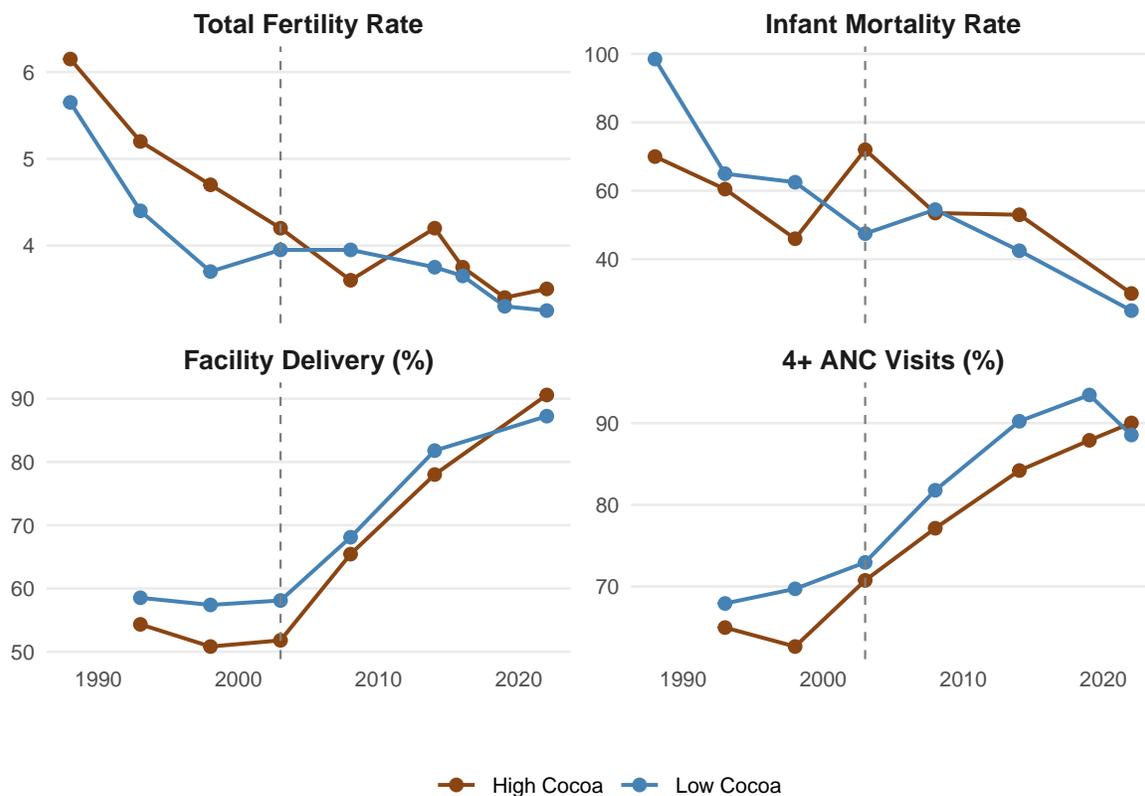


**Figure 5:** Outcomes by Cocoa Intensity: Forest Belt Regions, 1984–2010

*Notes:* Weighted means of key outcomes by census year and cocoa intensity (high cocoa: Western, Ashanti, Brong Ahafo, Eastern; low cocoa: Central, Volta). Forest belt regions only. Person weights applied. Source: IPUMS International, Ghana PHC.

## Extended Health Trends by Cocoa Intensity (DHS Data)

Dashed line = NHIS launch (2003)



**Figure 6:** Extended Health Trends by Cocoa Intensity (DHS Data, 1988–2022)

*Notes:* Regional means of DHS indicators by cocoa intensity. High cocoa: Western, Ashanti, Brong Ahafo, Eastern. Low cocoa: Central, Volta, Greater Accra, Northern, Upper East, Upper West. Dashed line = NHIS launch (2003). Source: DHS STATcompiler API.