Labor Migration, Capital Accumulation, and the Structure of Rural Labor Markets*

Taryn Dinkelman Grace Kumchulesi Martine Mariotti

October 2023

Abstract

Between 1967 and 1974, a bilateral treaty increased circular labor migration from Malawi to South Africa by 200%, bringing over 53 million USD in earnings into origin communities. A deadly migrant worker plane crash in 1974 ended these flows and led to migrant repatriation. We study how this shock affected local labor markets. In regions receiving more migrant capital after the crash, workers, particularly women, shifted from farming into non-farm work over thirty years. Investments in non-farm physical and human capital contribute to these sectoral changes. This natural experiment shows that temporary capital inflows can permanently reshape rural labor markets.

JEL Codes: J21, J24, O15, O16, O47, O55

*Dinkelman: University of Notre Dame, BREAD, CEPR, IZA, and NBER. Email: *tdinkelm@nd.edu*. Kumchulesi: Malawi National Planning Commission. Email: *gkumchulesi@npc.mw*. Mariotti: Australian National University and Stellenbosch University. Email: *martine.mariotti@anu.edu.au*. UK DFID and the IZA funded this work for the benefit of developing countries (GA-C2-RA4-181). Views expressed are those of the authors and not necessarily those of DFID (now FCDO) or IZA. We thank Malawi National Archives officials and Lucy McCann at Rhodes House Library for data collection assistance, and Freed Kumchulesi, Khwima Singini, and Regina Mannino for excellent research assistance. We also thank numerous seminar participants for constructive feedback.

What factors facilitate a structural shift of employment out of agriculture? What do workers in poor, rural areas require in order to move away from low productivity farming and towards more productive jobs in manufacturing and services? Such questions are at the heart of classic theories of economic development (Lewis, 1954; Rosenstein-Rodan, 1943) and continue to motivate economists to explore why sectoral productivity gaps persist, and what factors might trigger structural change in low-income countries.[1]

Theories of structural transformation typically rely on exogenous productivity growth or income shocks to generate incentives for labor to move across sectors. Recent empirical evidence points to specific technology and trade shocks as triggers for structural transformation in the labor market. For example, in Brazil, new labor-saving hybrid seeds released workers from low- productivity agriculture, shifting them towards industrial jobs (Bustos, Caprettini and Ponticelli, 2016; Bustos, Garber, and Ponticelli, 2020). Trade liberalization in Vietnam reallocated farm workers towards higher productivity non-farm enterprises by changing relative goods prices across these sectors (McCaig and Pavcnik, 2018). And in China, accession to the WTO reduced tariff uncertainty to a greater extent in the non-farm relative to the farm sector, resulting in employment shifts out of agriculture (Erten and Leight, 2021).

Our paper provides new empirical evidence on an alternative trigger for long-run structural change: a temporary increase in the local supply of capital provided by international migrant workers. Analyzing a natural experiment in Malawi that dramatically changed opportunities for international migration over a short time period, we show that a temporary migrant capital surge led to persistent changes in the structure of local labor markets that emerged over the following thirty years.[2]

[1]Gollin (2002) and Gollin, Lagakos and Waugh (2014) analyze productivity gaps across sectors. McMillan, Rodrik and Verduzco-Gallo (2014) document structural change in African countries since the 1990s.

[2]Dinkelman and Mariotti (2016) exploited features of this natural experiment to show how international migration shocks affect contemporaneous human capital attainment.

We exploit spatial variation in migrant money injected into local labor markets in Malawi between 1974 and 1975 following large, unanticipated shocks to international migration opportunities. In 1967, lucrative opportunities to migrate to the South African gold mines were opened up when a new treaty between the two countries facilitated legal, temporary migration contracts. This opportunity ended unexpectedly in the mid-1970s, with the banning and subsequent repatriation of all Malawian migrants in response to a mineworker plane crash that killed miners in transit in 1974. Migrants at different points in their time-limited contracts at the time of this plane crash were repatriated with different amounts of money, generating exogenous spatial variation in migrant capital per worker.

Between 1967 and 1975, legal temporary migration to South Africa rose by 200% and then fell to zero. This short, sharp shock to migration caused a large capital injection: Malawi received over 53 million USD in compulsory migrant remittances during this time; three times larger than US foreign aid to Malawi in 1974. Online Results Appendix Figure 1 shows this temporal variation in labor migration (on the left hand side y-axis) and migrant money (on the right hand side y-axis). One quarter of remittances between 1967 and 1977 occurred in 19 months following the plane crash. We hone in on 1974-1975 to leverage district-level variation in migrant capital generated by the exogenously-timed plane crash.

To study how these migrant capital injections affected local labor markets, we build a panel dataset of money flows and labor market outcomes by digitizing archival material on district-level remittances and matching it to multiple waves of Malawian population census data. The main treatment variable is the amount of migrant capital per worker received by each district immediately after the migration ban, when all migrants who were still away had returned home, and were paid out their accumulated deferred pay.

Our empirical strategy compares subsequent changes in labor market outcomes from 1977 onwards across districts that receive more versus less migrant money per returned worker, controlling for district and decade fixed effects, and allowing for time-varying effects of baseline covariates. The key identification assumption is that conditional on district and decade fixed effects, and controlling for differential trends related to baseline covariates, districts receiving more or less capital per worker after the plane crash would have evolved similarly in the absence of this injection of capital.

We show that districts with larger capital injections per migrant experienced a greater share of workers shifting out of farming and into non-farm service activities after the end of the international migration episode. Sectoral shifts began slowly in the first ten years following the end of migration, once all migrants had returned, and persisted into the second and third decades post-shock, long after the migrant capital surge dissipated. Women and men shifted into the service sector, specifically into construction (men) and into retail sector jobs (both). Impacts are larger, and most robust, for women.

These shifts in sectoral employment are unlikely to be driven by miners acquiring more education or changing their aspirations while abroad. Mine work was extremely low-skilled, and miners were constrained to live in mining compounds, physically segregated from the South African economy during apartheid. Moreover, our results are strongest in subsequent decades for women, who would not have been migrant mineworkers at all.

The biggest threat to identification is the possibility that the structure of work in high migrant capital per worker districts might have changed even in the absence of the temporary migrant capital infusion. Because disaggregated employment data do not exist for Malawi prior to the 1970s, we cannot directly test for parallel pre-trends in employment outcomes across high and low migrant capital per worker districts. Instead, we test for parallel trends in population growth – another typical measure capturing the level of development in a district – prior to the plane crash. We cannot reject parallel trends in population growth prior to 1967, suggesting that high and low migrant capital per worker districts were not on differential growth paths before the migrant capital injection. However, we show that the migrant capital injection triggers population growth across districts after the migration episode, as well as a spatial shift in population towards market towns and urban areas within a district. These changes in the distribution of population within a district are additional indicators of structural change in the rural economy.

The evidence in our paper is consistent with migration-induced capital accumulation triggering a slow and steady expansion of the non-farm service sector in rural Malawi. Economic theory suggests this could have happened through some combination of a temporary positive income shock and an increased supply of capital invested in the farm and/or non-farm sectors (Herrendorf, Rogerson and Valentinyi, 2014).

To investigate mechanisms for the persistence of changes in the labor market, we assemble data on farm and non-farm capital goods from population and agricultural censuses, household income and expenditure surveys, and demographic and health surveys, before and after the migration shock. While our data are unable to tease out all possible mechanisms of influence, we provide evidence that both human capital accumulation and non-farm physical capital (rather than farm capital) likely account for some of the long-run changes we identify. Human capital gains in high migrant capital districts are larger for females than for males. And, fifteen years after the end of the temporary migration episode, households in districts that received more migrant capital were significantly wealthier and held more private assets. A key result we show is that these wealth gains are not solely driven by younger cohorts, who would have benefited from human capital gains funded by migrant earnings. Instead, older cohorts also benefit from being in higher migrant-capital districts, suggesting that education gains are only a part of the story behind the structural changes we measure in our data.

The primary contribution of our paper is to empirically demonstrate that a once-off capital shock, facilitated by temporary international migration, can trigger persistent changes in labor market structure and shifts out of agricultural work over the long-run. Our paper contributes to the literature on the broader impacts of remittances in developing countries.[3] Clemens and McKenzie (2018) note that the cross-country literature on effects of remittances

[3]There is a long tradition of estimating how migrant money directly affects migrant households, e.g. Dustmann and Kirchkamp (2002), Woodruff and Zeneto (2007), Yang (2008), Kinnan, Wang and Wang (2016).

on home country outcomes is largely inconclusive, and that economists remain *"surprisingly unsure of its (remittances) broad development effects"*. Addressing this deficit in the literature, a handful of recent empirical studies use natural experiments and exogenous shocks to migrant demand or to remittances *within* countries to uncover the broader, market-level development impacts of international migration. For example, Theoharides (2018) shows that increased (exogenous) demand for migrants from the Philippines stimulates higher levels of schooling attainment at local labor market level. Theoharides (2020) also finds that shutting down the largest Filipino migration channel to Japan reduced income, and increased labor force participation and child labor in sending labor markets of the Philippines. These two papers demonstrate that shocks transmitted through international migration can have profound effects on sending country economies.

Our paper is most closely related to Khanna, Murathanoglu, Theoharides, and Yang (2022), who show that an exchange rate-induced shock to migrant remittances to the Philippines has positive long-run impacts on local incomes in sending labor markets. However, where their findings are based in a setting with long-standing and continued migration flows and focus on outcomes two decades after an exchange rate shock, we show that even once-off episodes of temporary migration can have long-lasting (up to three decades), persistent effects in the sending economy. Both our study, and theirs, demonstrate that migrant capital matters for long-run outcomes in local labor markets, and that decades of data are required to detect impacts on employment and on incomes.

A second key contribution of our paper is to highlight the potential importance of temporary international migration and associated migrant capital for Africa, the region with perhaps the greatest potential for structural change. Migration is an important feature of labor markets in many low-income countries. Remittance flows are now the largest source of external funds in developing countries, exceeding total foreign aid flows and foreign direct investment (The World Bank, 2019). Only a handful of papers estimate the developmental impacts of seasonal and guest worker programs on origin countries, all of them outside of Africa, e.g. Gibson, McKenzie and Rohorua (2014) and Gibson and McKenzie (2014). Malawi's experience suggests that managed circular labor migration channelling earnings back to sending communities may offer a practical tool for triggering longer-run structural change in communities where industrial, agricultural, and trade revolutions have been slow to arrive. There is vast scope to learn more about how these tools should work in practice. For example, the changes we measure in Malawian labor markets occur after the end of international migration. Our setting is not well designed to tell us whether the sectoral employment shifts in response to migrant money would have been even larger, or would not have occurred at all, if international migration had been allowed to continue. This remains a set of questions for future work.

Finally, our results provide new evidence on the triggers of structural transformation. Similar to Bustos, Garber, and Ponticelli (2020), we highlight the role that an increased supply of capital plays in facilitating structural change, although the increase in capital here is transitory, being driven by an increase in earnings from temporary labor migration (not from agricultural technology shocks as in Bustos, Garber, and Ponticelli (2020)), and we measure impacts over a longer horizon. To the best of our knowledge, the structural change literature has not yet considered international migration as a source of local capital that could facilitate reallocation of labor across sectors.[4]

1 Conceptual framework: Linking temporary capital injections with long-run changes in the labor market

How might a temporary injection of migrant money affect the types of work that people are able to do in a rural economy? First, and most directly, the return of migrant money increases local incomes and consumption of migrants and migrant households. Returning

[4]Foster and Rosenzweig (2008) model how national rural-to-urban labor migration and remittances facilitate structural change in rural labor markets.

migrants increase their local demand for farm- and non-farm goods. When preferences are non-homothetic, the demand for – and hence viability of – goods and services produced by the non-farm sector will increase faster than the demand for food (Herrendorf, Rogerson and Valentinyi, 2014; Gollin, Jedwab and Vollrath, 2016). Returns to work change across sectors, bidding up the price of non-farm relative to farm work and encouraging workers to move out of farming. Yet these changes are likely to be transitory, unless there are mechanisms for capital accumulation to occur in response to the temporary increase in migrant capital. This is where the following channels of impact become important.

A second way that migrant capital may affect local labor markets is by providing liquidity for investment in all sectors and across a broader range of households. In the farm sector, farmers may choose to invest in farm capital such as seeds, fertilizer, or farming equipment. If the economy is closed, as is the case for Malawian districts, more farm capital lets farmers meet minimum food production requirements with less labor, and excess labor can be released to the non-farm sector.[5]

In the non-farm sector, migrant capital can enable individuals to overcome fixed start-up costs of opening businesses, and/or allow them to expand production in this higher productivity sector (as in Woodruff and Zeneto (2007) in Mexico, and Yang (2008) in the Philippines). Even if migrants are not starting businesses themselves, they may act as local financial intermediaries that lend money to entrepreneurs (Buera, Kaboski and Shin, 2013), thereby allowing the non-farm sector to expand and pulling workers out of the farm sector.[6]

Investment of migrant earnings in human capital (education or health) is a third channel through which migrant capital can affect local economies. Investments in health or education

[5] There is wide variation in maize prices across districts in Malawi (Robinson, 2016), indicating that districts are reasonably closed economies. Researchers (e.g. Beegle, Galasso and Goldberg (2017)) have also shown large frictions in land and in agricultural input markets.

[6] Most borrowing is likely to be local in scope. As late as 2014 (The World Bank, 2014), Malawi had only 4.85 ATMS and 3.2 bank branches per 100,000 people respectively.

feed into the labor market with a lag, directly improving worker productivity of the next generation. If literate or healthier workers have higher returns in non-farm relative to farm sectors, an increase in the share of skilled workers could lead to shifts in the sectoral allocation of labor over time (e.g. as in Porzio, Rossi, and Santangelo (2021)).

Underlying each of these investment channels is the idea that there is insufficient capital in the local economy to break out of a poverty trap (Balboni, Bandiera, Burgess, Ghatak, Heil , 2021). Incomes are so low that savings cannot be sustained, and people cannot afford to buy much more than food. In this context, increasing the supply of capital even temporarily can lead to a big push through some combination of the above channels: some households spend more on non-food items (services, physical capital, human capital); other households start up small businesses to supply this new demand. Importantly, in order for temporary increases in migrant capital to trigger long-term impacts on the structure of the labor market, the shock to migrant capital should translate into higher savings and investments at the level of the local labor market, not just at the level of the individual migrant. As savings and investment accumulate within and across migrant households, the impact of the initial income shock grows and becomes visible at district level over time.

In Section 6, we will assemble evidence on these channels of persistence and discuss why migrant capital injections may be slow to effect change at the level of the local labor market.

2 Context: Rural labor markets in Malawi

Malawi was and still is a predominantly agricultural economy, although the importance of agriculture for work and output has changed over time. In the 1960s, agriculture accounted for half of Malawian GDP. By 2015, this contribution had shrunk to 30%. Employment in agriculture declined much more slowly from 84% in 1977 to 61% by 2008. The share of manufacturing in GDP fell from an already low 13% in the 1960s to 10% in 2015, with employment in this sector shifting from 5% in 1977 to 9% in 2008. Over five decades, the share of services in GDP increased from around 40% to almost 55% (The World Bank, 2016) with the share of employment in the sector rising three-fold from 8% in 1977 to 24% in 2008. These shifting patterns of employment from farm to non-farm sector, with non-farm jobs concentrated in services rather than manufacturing, strongly resemble patterns of structural change in other African labor markets over the last three decades (e.g. Gollin, Jedwab and Vollrath (2016) and McMillan, Rodrik and Verduzco-Gallo (2014)).

Online Results Appendix Table 1 illustrates these broad patterns of change for men and women separately. Between 1977 and 2008, the share of workers in the Malawian agricultural sector fell from 94% to 70% for women and from 76% to 53% for men. Over the same period, the share working in services increased dramatically. In 1977, 2.8% of women and 12% of men were in service sector jobs and by 2008, these shares had risen to 20% for women, and 28% for men. A Herfindahl index of sectoral employment concentration shows a fall from 0.89 to 0.55 for women's employment and from 0.62 to 0.36 for men's employment, signifying a decline in employment concentration over time.

It is worth noting that labor force participation rates for men and women are high in Malawi. Between 84% and 96% of working adult adults were working or looking for work in different decades. At the same time, there is good evidence of substantial underemployment and excess labor supply in rural Malawi (see, for example, Goldberg (2016)).

The non-farm sector in rural Malawi is dominated by work in retail trade, the public sector, and small-scale personal and general services, often run by the self-employed. To fix ideas about the nature of this non-farm work, we graph the percentages of workers employed in non-farm industries and occupations, for each of the most prevalent non-farm categories of work. Online Results Appendix Figure 2a shows that retail and wholesale trade tops the list of non-farm industries: 45% of women and 32% of men work in this sector, with the next largest categories being the public sector (e.g. teachers, medical staff) and hotels and restaurants for women, and manufacturing, construction, and the public sector (including defense) for men. Smaller shares of women work in manufacturing, transport and communications, and other services, and 6% of men work in business services or transportation/communications/storage. Occupational patterns in Online Results Appendix Figure 2b follow the industry patterns in the prior figure. Almost 30% of women working in the non-farm sector are working proprietors (self-employed). 17% are sales and shop assistants, 16% are teachers or nurses, and 6% are food and beverage producers. About 18% of men report being working proprietors and 11% are shop assistants; 19% are teachers, nurses, or work in security while 12% work as brick-makers or carpenters.

We note several facts about how services are produced in Malawi. First, small businesses are common in rural areas. While all rural households farm, data from the mid-1990s shows that around one in five households also owns a non-farm enterprise (Alvarez-Cuadrado, Amodio, and Poschke, 2021). Second, although many of the service sector jobs documented in Online Results Appendix Figures 2a and 2b are in small-scale household enterprises, these activities tend to use more capital on average than farming does.[7] Third, labor tends to be more productive on average in the non-farm sector than in farming: estimates suggest that labor is two to four times as productive in the non-farm relative to the farm sector (Gollin, Lagakos and Waugh, 2014).[8] So, even though jobs in the non-farm sector are low-skilled and not highly productive, the average worker in this sector may still be more productive in such a job relative to working the farm sector.

With these stylized facts about Malawi's rural economy in hand, we next describe the historical context of our natural experiment.

[7]Farming uses little capital in Malawi (Chen, Restuccia, and Santaeulalia-Llopis, 2017).

[8]Gollin, Lagakos and Waugh (2014) use micro data from developing country household surveys to compute average labor productivity gaps between agricultural and non-agricultural sectors of around 4.3 (for countries similar to Malawi, they find gaps of 2.3 for Ghana and 4 for the Ivory Coast). This means that shifting one worker from the farm to the non-farm sector should generate 4.3 times more value. We use income and expenditure survey data from Malawi to estimate the value added of labor in farm and non-farm businesses in rural areas; we find a productivity gap of 2.9 in 2000.

3 Historical circumstances of the natural experiment

Off and on during the 20th century, mine work in South Africa provided a feasible sector of work for Malawian men willing to migrate. These work opportunities were organized between the Malawian government, whether colonial or independent, and the South African mine labor recruiting agency, the Witwatersrand Native Labour Association (*Wenela*). Workers were employed on fixed two-year contracts at set rates that were higher than any local wage work, which itself was limited. Selection into mine work on the basis of observable characteristics was limited to minimum age and weight requirements. There were no education or skill requirements and about 60% of the Malawian miners in South Africa had no education at all. Most miners were between the ages of 20 and 30 (author calculations, South African Census data 1970 and 1980). Prior to 1967, legal circular migration was limited by national recruiting quotas that were held at less than 2% of the working age male population.

Then in 1967, a new labor treaty (Treaty, 1967) removed the old quota and migration expanded from 40,000 to over 120,000 men in five years, as seen in Online Results Appendix Figure 1. This short-lived surge in migration ended abruptly in April 1974, when a *Wenela* plane returning to Malawi crashed, killing all miners aboard. Then-president Banda banned all labor recruiting in Malawi and recalled all miners from South Africa. The number of Malawians working on South African mines fell to zero in the two years following 1974. By 1977, all miners were back home in Malawi.[9]

Between the labor treaty (1967) and the labor ban (April 1974), mine work expanded dramatically across all districts, facilitated by low entry barriers and wages that promised 2.5

[9]By 1977, Banda had realized that mining capital was a crucial source of foreign reserves for the country and rescinded the ban on migration. Migrant flows never returned to prior levels; by the 1980s, *Wenela* had redirected recruiting towards the South African labor market. Mariotti (2015) analyzes how this shock affected outcomes in South Africa.

times more than the next best wage-earning job at home.[10] By the late 1970s, approximately one in three Malawian adult men had ever worked as an international migrant (see summary statistics in Online Results Appendix Table 2). Crucially for our paper, because these miners were contractually obliged to receive 60% of their earnings as deferred pay upon repatriation, most migrant earnings returned to Malawi. This deferred pay formed the basis of their ability to accumulate in rural sending regions.[11]

Flows of money paralleled the ramp up of migration in the late 1960s and the equally dramatic decline of migration in the mid-1970s. Between 1966 and 1975, total migrant capital flows rose twenty-fold, driven by increases in migration and by increases in the mining wage offered in South Africa (see Online Results Appendix Figure 1; deferred pay in June 1966 was 135,604 USD; by November 1974 it had risen to 2,786,807 USD).[12] Total deferred pay inflows over the entire period were 53 million USD. At peak migration in early 1973, Malawi received 2.75 million USD from miner earnings in a single month, or almost 115,000 USD on average per district. Each migrant returning from a completed two-year contract would have received between 130 USD and 295 USD, depending on when he left. As a benchmark, average per capita GDP in 1970 in Malawi was only 63 USD.

Money flows spiked after the plane crash as all miners were repatriated (see second vertical line in Online Results Appendix Figure 1). This spike lasted just under two years. Depending on how one discounts, between one fifth and one quarter of the total amount of migrant capital from the entire migration episode returned to Malawi in the 19 months after the plane crash. This latter part of the migration period (1974-1975) represents the

[10] Wage work for farmers was most prevalent on tobacco or sugarcane estates.

[11] We refer the reader to Dinkelman and Mariotti (2016) for a more detailed discussion of the history of organized legal mine migration from Malawi to South Africa.

[12] Figure 3 in the Online Results Appendix shows the trend in miner real and nominal wage rates in Malawian Kwacha between 1966 and 1975 alongside the global gold price in USD, the key driver of increased wages.

period of largest, coordinated capital flows back to rural districts in Malawi. Because not all miners would have completed their two-year contract at the time of the plane crash, the average amount of money per migrant returned to the country after April 1974 was 73 USD per migrant.

We are concerned with estimating the impact of migration-induced remittances from this temporary migration episode on district-level changes in the labor market. Capital flows per worker throughout the Treaty and ban period fluctuated across districts, we argue, somewhat randomly. The total amount of capital per migrant that flowed back to a district was driven by several factors: the number of migrants contracting at specific wage rates (with higher wages being paid later in the period) and, at the end of the Treaty period, by how much of his contract each miner had completed upon repatriation (with those having completed more of their contracts having more paid out). This variation in capital per returned migrant, driven by timing differences in contracts, is what we exploit to identify effects on sectoral employment shares in the labor market. The next section discusses alternative ways of measuring this capital inflow.

4 Empirical strategy and Data

4.1 Estimation

To isolate the persistent effects of migrant capital per worker at market-level, we specify the following empirical model for labor market outcomes Y_{dt}:

$$Y_{dt} = \sum_t \alpha_t MigrantCapital_d * Decade_t + \kappa_t + \delta_d + W_d Trend_t \lambda + \epsilon_{dt} \qquad (1)$$

where d is the district and t is the decade (1977, 1987, 1998 or 2008). The main set of outcomes Y_{dt} includes the share of economically active adults in agriculture, manufacturing or services. $MigrantCapital_d$ is our main treatment variable which captures migrant capital per returning worker; we discuss measurement of this variable below. $MigrantCapital_d$ is interacted with $Decade_t$, a set of decade dummies for one, two and three decades after the end of migration (1987, 1998 and 2008 respectively). The omitted decade category, 1977 is our baseline year and the year by which all migrants have returned from South Africa.

Our regressions include decade (κ_t) and district (δ_d) fixed effects, and a vector of baseline district covariates and region dummies interacted with a linear trend term ($W_d Trend_t$). ϵ_{dt} is an idiosyncratic error term. Regressions are estimated separately by gender.

Equation (1) allows us to estimate the effect of the capital per returned migrant on differential changes (rather than level differences) in employment outcomes across districts after 1977. δ_d controls for constant average differences in labor markets across districts (e.g. districts with lake access are always able to support work in fishing industries). These controls also standardize for district population size. κ_t controls for aggregate changes in the labor market that affect all workers equally, for example, a nationwide drought that occurred in the early 1990s and which likely affected agriculture everywhere. Trend interactions flexibly allow districts with different initial population densities, literacy rates, marriage rates, malaria risk, cash wage work, and crop suitability for tobacco and maize farming to evolve differently over time. Regions are also allowed to evolve differently over time, by including linear trend interactions with Central and Southern region indicators (Northern Region is the omitted category).

Both sign and significance of the α_t parameters are informative. α_{1987} is the percentage point change in the relevant employment outcome (e.g. share in farming) between 1977 and 1987, the first decade after the plane crash. α_{1998} and α_{2008} provide the same parameter estimates for two and three decades post-crash. For the migrant capital injection to have had any effect on the local economy, we should consider α_{1987}. Non-zero estimates of α_{1998} and α_{2008} indicate persistence of the initial shock over time.

4.2 Identification

Our identification assumption is that conditional on district and decade fixed effects, and controlling for differential trends related to baseline observable characteristics, districts receiving more or less capital per worker would have evolved similarly in the absence of this injection of capital.

We consider two ways to measure the migrant capital injection. One approach is to use cross-district variation in migrant capital per worker received over the whole migration period, from 1967 to 1975: we call this total *K/L*. Another approach is to hone in on the 1974-1975 period to leverage variation in migrant capital per worker across districts generated by the exogenously-timed plane crash. We call this measure *K/L shock*.

Each measure has its strengths and weaknesses. Total *K/L* is attractive because it could capture the effects of cumulative capital in a district. This measure could be important for detecting agglomeration effects that contribute to labor market adjustments over time. Yet, as we noted in Section 3, some of the variation in total *K/L* is a function of when a worker left for the mines. If we are worried that unobserved local economic factors drive the time profiles of migration across districts over this eight year period, we would want to test that trends in sectoral employment shares before 1977 are similar across high and low total *K/L* districts. Unfortunately, the data for this pre-trends test does not exist.[13]

Alternatively, we could use the the cross-district variation in migrant capital per worker induced by the timing of the plane crash. *K/L shock* is appealing because the plane crash is clearly an exogenous shock and allows us to pinpoint a much finer source of variation in migrant capital per worker in each district. Remittances triggered as a result of the plane crash would vary at district-level only because workers contracted at different wages over a much smaller window (whether in 1972, 73 or 74) and started at only slightly different times, leading to different total months of contract completion at the time of the crash.[14] The

[13]Census instruments from the 1940s, 1950s, and 1960s did not collect sector of work data (nationally or at district-level) for any African workers.

[14]For example, a district with more migrants closer to the end of their 24 month work period in April 1974 would have received more money; a district with more workers who had started slightly later would also have received more money because of the higher wage.

differences in the composition of these contracts around the time of the plane crash is likely to be random. However, using this measure means that we might miss out on estimating the effects of the cumulative migrant capital receipts.

Because of the data limitations we face in checking historical pre-trends in outcomes variables in the context of Malawi, we make the "cleaner" variation in migrant capital, *K/L shock*, the basis of our identification strategy and present results using this measure of treatment. However, we also show that our main effects on sector of work are similar using the broader total *K/L* measure.

4.3 Evidence to support the identification strategy

Table 1 presents tests of covariate balance across high and low *K/L shock* districts. For comparison, we also show the balance regression using the alternative treatment variable, total *K/L*. *K/L shock* is contained in *K/L*, and both outcomes are measured in units of 100USD per migrant worker. To gauge the economic significance of the point estimates, coefficients should be multiplied by 100 (USD).

We regress each district-level outcome on a set of historical and geographic district characteristics. Since the cross-sectional variation comes from 24 districts, we report randomization inference p-values for each coefficient as well as for the joint test of all slopes equal (Roodman et al, 2019; Young, 2019; Davidson and MacKinnon, 2010).

Comparing the results across the two columns, we note that the coefficients in the *K/L shock* regression are always smaller than the coefficients in the *K/L* regression, and in some cases, by a large margin. At the same time, none of the coefficients is statistically significant in either regression, and we fail to reject the joint test of significance in either case. Since districts look statistically similar across high versus low *K/L Shock* and *K/L* districts, and because we have argued that the cleaner variation is contained in *K/L Shock*, the bulk of our analysis will be focused on using *K/L Shock* as the main treatment variable.

Some coefficients in Table 1 (such as those on the variables capturing the share of men and women working for no cash wages) are large in magnitude. For example, a district in the Central region is predicted to receive 47USD more in *K/L Shock* than a district in the Northern region (omitted category). To interpret the magnitudes of continuous covariates, we multiply the coefficient by the value of one standard deviation of the relevant variable (see Online Results Appendix Table 2 for covariate means and standard deviations) and then multiply by 100(USD). With a one standard deviation higher share of men earning no cash wage in 1966 (0.1), a district is predicted to receive 179USD (-17.97*0.1*100) less per migrant after the crash. With a one standard deviation higher share of women earning no cash wage (0.14) in 1966, a district is predicted to receive an additional 158USD per migrant (11.29*0.14*100) in the wake of the crash. For other continuous covariates, the correlation with migrant capital measures is very small: for example, a district with a one standard deviation higher value of maize suitability is predicted to receive only 8.5USD less from the *K/L Shock* (0.14*0.006*100).

While it would be ideal to have small correlations between all baseline covariates and the migrant capital measures, our empirical strategy does not require that districts be balanced in the cross-section. Equation (1) allows districts with different values of baseline variables to trend differently over time, and we will show that our results are robust to including or excluding the trend controls in $W_d Trend_t$.

Although we cannot test for pre-trends in employment outcomes due to data limitations, we can examine what happens to log population and urbanization rates within districts before and after the capital shock. Testing for pre-trends in population outcomes is a reasonable alternative since population density and urbanization are typical proxies for economic development. Because we can measure these outcomes back to 1945 (population density) and to 1966 (for urban share of the district), we can estimate Equation (1) using a more standard difference-in-differences design. The results from these analyses show no evidence of differential pre-trends across high and low *K/L shock* districts.

Because we are estimating the impacts of a shock over the thirty years following the initial event, one might be concerned about a variety of potential threats to validity that could arise over this long time horizon. To counter the most direct and obvious potential threats, we verify the robustness of our results in several ways.

A key potential threat to validity is if some other shock to local labor markets occurs in high (or low) *K/L shock* districts in years after the plane crash. Because agriculture is such a large component of Malawi's economy, candidate shocks are likely to have come from that sector. Our reading of the economic history literature suggests two candidate confounders: the tobacco liberalization program which occurred in the early 1990s, allowing subsistence farmers to start producing tobacco for export, and national fertilizer subsidy programs introduced in the late 1990s and mid-2000s.

To address potential confounds coming from agricultural sector policies that differentially affect districts, all main results include tobacco suitability and maize suitability interacted with a linear trend in $W_d Trend_t$. To test for potential confounding from agricultural input subsidies distributed in the mid-2000s, we additionally control for voucher access in the mid-2000s interacted with the 2008 decade dummy (results are robust). Our results are also robust to including nonlinear controls of all covariates interacted with decade dummies, to including region-decade fixed effects, and to scaling the migrant capital shock by baseline population in 1966, rather than by the number of returning migrants.

4.4 Data and summary statistics

We measure labor market outcomes, population outcomes, and district-level covariates using six waves of Census data between 1945 and 2008 (different outcomes are available in different sets of years). We digitize historical Census data available at the district-gender level from 1945, 1966 and 1977 and match this with Census labor market data from 1977 (100% sample), 1987 (the 10% sample), 1998 (100% sample) and 2008 (10% sample). Details of variable construction are in the Online Data Appendix.

Key labor market variables are defined for men and women using labor market questions that are generally the same across survey instruments. Broad industry of work is available for all economically active individuals 10 years and above. We use these broad measures of industry – agricultural, manufacturing and services – to look at labor reallocation across sectors, as well as a finer breakdown of industries in the non-farm sector: general manufacturing, retail, transport and communication, and all other services, including personal services and government employment. Using the broader measures of industrial sector, we construct a Herfindahl index that captures the diversity of employment within the district. Data on total population and population by gender are available for each district from 1945 onwards and on urban shares of population from 1966 onwards.

We also collect and digitize material from *Wenela* administrative records showing district-month level remittance flows from October 1967 to November 1975. Three categories of monies were recorded by the mine labor recruiting agency: deferred pay, remittances and other deposits. 89% of all monies returning to Malawi were in the form of compulsory deferred pay; this is the measure of migrant capital we use in our analysis. Amounts are aggregated over time to the district-level.

We estimate the number of migrants returning to each district in different years by combining information from two variables in the 1977 Census data. We estimate the district-specific number of migrants for the entire migration period (1967-1977) by multiplying the district-specific number of men who report ever migrating at all in the 1977 Census by the national share of these ever-migrants who report returning to the country between 1967 and 1977. To estimate the number of migrants returning to each district after the plane crash, between 1974 and 1977, we make a similar calculation: multiplying the district-specific number of men who report ever migrating at all in the 1977 Census by the national share of these ever migrants who report returning to the country between 1974 and 1977. Online Results Appendix Figure 1 shows that the number of returnees peaked in 1974 and 1975, just after the plane crash.

Means and standard deviations for the main covariates in our analysis dataset are in

Online Appendix Table A.2. An important takeaway from this table is that international migration was widespread in Malawi prior to the plane crash. By 1977, on average one in three adult men in a district had worked abroad. Between 3 and 9% of men in each district were actually absent from Malawi at the time of the plane crash.

The average district received 87.6 Kwacha (the equivalent of 73 USD) per migrant worker after the plane crash; this is the mean of *K/L shock*, representing more than one year's worth of earnings (1970 GDP per capita was 63USD). Each *person* in an average district would have received 7.8USD between May 1974 and November 1975, or about a 12% increase in income relative to per capita GDP.

5 Main results

5.1 Impacts of *K/L shock* on broad sector of work

Table 2 presents our main results for the broad categories of employment in agriculture, manufacturing and services, and the Herfindahl measure of employment diversification. Panel A shows estimates for women, Panel B for men. The unit of observation in each regression is the district-gender-decade. For each outcome, we present estimates of α_t from Equation (1), including all controls. Regressions are weighted using 1977 population weights. Following Young (2019), who argues that p-values generated through randomization inference are more reliable than inference based on robust clustered standard errors in situations with few clusters, we report statistical significance showing p-values generated using randomization inference in square brackets. In discussing results, we evaluate each coefficient at the mean amount of capital per returned migrant for the post-plane crash period (multiplying $\hat{\alpha}$ by 0.73, or 73 USD per migrant).

Table 2 shows that districts with larger capital flows per returned migrant experienced more structural change away from agriculture into the non-farm sector, with both women and (to a lesser extent) men shifting into manufacturing and services and with some increase in the diversity of employment (reduction in the industry concentration of jobs).

Specifically, for each additional 73 USD per worker that flowed back to a district by 1977, the share of women working in agriculture fell by 0.65 percentage points in the first decade following the shock (-0.009*0.73*100), by 2.7 percentage points in the next decade, and by 3.2 percentage points by the third decade after the shock. Column (2) shows smaller shifts of women into manufacturing (0.16 to 0.38 percentage points) and column (3) shows larger shifts into services (0.3 to 3.57 percentage points by three decades after the shock). Diversity in female employment also increases (the industry concentration index falls). Randomization inference p-values indicate that these sectoral shifts for women are precisely estimated for agricultural and service sectors, and for manufacturing employment and industry concentration in the second and third decades post plane crash. Point estimates are most precisely estimated in the very first decade after the end of the migration episode.

Panel B of Table 2 shows that men experience similar but smaller shifts away from agriculture and towards non-farm sectors. At the mean value of migrant capital per worker, the share of men working in agriculture fell by 0.8-1.4 percentage points in the first two decades after the shock and rose by 0.16 percentage points in the third decade. Shifts into the manufacturing and services sectors were also positive, but small. Male employment in services increased by a significant 1.3 percentage points in 1998 and a smaller 0.28 and 0.45 percentage points in the first and third decades. Shifts into manufacturing are at most 0.35 percentage points (and statistically significant) in the second decade post-crash.

Although point estimates for women are always larger than those for men, these differences are not always statistically significant using randomization inference p-values. In Panel C of Table 2, we show that male and female impacts are statistically significant only for shifts out of agriculture in the first decade post-crash, the shift into manufacturing in the second decade post-crash, and the industrial concentration index in the first decade post-crash.[15]

[15]We pool the male and female datasets, generate a full set of female interactions for every covariate, re-estimate Equation (1) including this full set of interactions and test the significance of the $K/L\ shock * Decade_t * Female$ term for each decade using randomization

To be sure that results using our main treatment measure *K/L shock* are not missing important agglomeration economies, we estimate Equation (1) using the alternative treatment measure of total migrant capital per worker returned over 1967-1975. We should expect to see cumulative migrant capital matter in the same way for sectoral labor reallocation, or even more so, using total migrant capital per worker. The mean of the treatment variable here is 1.29. Table 3 results indicate that at the mean K/L in a district, female employment shifts out of agriculture by 0.68, 2.7, and 3.22 percentage points in the three decades after 1977, and shifts female employment into services by 0.36, 1.81, and 3.61 percentage points. These results are very similar in magnitude and significance to the results in Table 2. Men's results are also similar across alternative treatment measures: at the mean level of K/L, men's employment in agriculture falls by 0.9 and 1.5 percentage points in the first two decades post-crash, and men's employment in services rises between 0.28 and 1.4 percentage points. The similarity of impacts across the two measures of treatment suggests that we are not missing scale economy effects by using the "cleaner" variation in *K/L shock*.

Overall, employment shifts in response to capital inflows are positive, somewhat lagged, and persistent. They suggest some measure of structural change: in districts with more migrant capital per worker, men and women shift out of farming and into services, and to a lesser extent, manufacturing.[16] We discuss the lagged and persistent nature of the employment impacts in Section 6.

We can use Table 2 results in a back-of-the envelope estimate of how much migrant capital contributed to overall structural change over time. In the four decades before 2008, employment shares in agriculture fell by around 24 percentage points for women, from 94% to 70% (Online Results Appendix Table 1). For men, this decline was 22.8 percentage points. Summing up our estimates of the percentage points from Table 2, the average district-level

inference. This is a demanding specification, given our 24 districts.

[16]Our main results are robust to a variety of different specifications, shown in Online Appendix Tables A.4-A.8.

migrant capital shock accounted for about 25% of the structural reallocation of female labor out of farming and into non-farm work. For men, this contribution is smaller, at around 5.8%.[17]

5.2 Impacts of *K/L shock* on narrow sector of work

What types of non-farm work developed in high capital shock districts, in the thirty years after this migration episode? Table 4 presents a finer breakdown of sector of work for the non-farm sector: general manufacturing and construction, general services, retail, and transport and communications. General services include personal services (e.g. guards, domestic workers and cooks), business services (advertising, or insurance, banks and engineers, legal services, accountants) and other services (e.g. barbers, tailors, typists, public sector workers). The retail sector includes wholesale and retail trade of food, fuel and other goods, hotels and restaurants, car repairs etc. Transport includes transport of goods and/or people, including using buses, taxis, boats, bikes, warehousing, and telecommunications. All specifications follow the form of Equation (1).

For women, we see the most significant and sustained impacts on employment in the general services and retail sectors. For each 73 USD per returned migrant received, the share of women working in general services rose between 0.2 and 1 percentage points; the share of women working in retail rose between 0.14 and 2.48 percentage points. The joint test for the shift of women into retail can reject zero at conventional levels of significance. The share of women working in each of the other non-farm sectors also rose significantly in every decade after the capital injection.

The shifts of employment within the non-farm sectors are a little different for men. Table

[17]An average district had 58,000 women and 57,000 men. The percentage increases in employment in services estimated in Table 2 translate into a 5.78 percentage point increase in employment among 58,000 women, or 3,353 more jobs. For men, the equivalent impact a 2.07 percentage point increase in service sector employment, or about 1,180 jobs.

4 Panel B indicates that more capital in the district shifted male jobs out of manufacturing (0.28 - 0.56 percentage points), transport (0.08 - 0.39 percentage points), and to some extent general services, while shifting male jobs into construction (0.36 - 0.95 percentage points) and retail (0.39 - 1.6 percentage points). The male employment shifts seem to be largest in the second decade post-shock, while the female employment shifts are largest in the third decade post-shock.

5.3 Impacts of *K/L shock* on population growth and urbanization

Next, we investigate how migrant capital inflows affected population growth and urbanization within districts. We estimate versions of Equation (1) using a set of population variables P_{dt} measured at district-year-age group level. We group the population into children (under age 5), youth (ages 5-18), and adults (ages 19 and older). For urbanization outcomes, the unit of observation is the district-year cell. t now includes six years of Census data from 1945 to 2008 for population outcomes and five years of data (excluding 1945) for the urbanization measure. Regressions are unweighted, and all controls (except population density for the population outcomes) are included. Randomization inference p-values are presented for coefficent estimates and for joint tests.

Figures 1a and 1b plot estimates of α_t, the relationship between the amount of migrant capital per worker received by each district between 1974 and 1975 and district-level population growth and urbanization outcomes before and after the migration surge. The omitted category for population outcomes is 1945, and 1966 for the urbanization outcomes. Each point on the line represents the marginal impact of receiving 73 USD of migrant capital between 1974 and 1975 on the log of the age-group specific population counts in the district in each Census year (Figure 1a) or on the share of the population living in urban areas in the district (Figure 1b). The regression table underlying these figures appears in Online Appendix Table A.3.

The figure shows that districts that were going to receive 73 USD in migrant capital between 1974 and 1975 did not look significantly different (population-wise) compared to districts that were about to receive less migrant capital, prior to 1966. This suggests that high and low migrant capital per worker districts were not on differential growth paths before the migrant capital injection.

Between 1966 and 1977, this pattern diverges. Districts receiving more migrant capital per worker by 1977 experience significant population growth by 1977. This growth occurs immediately, is sustained over the next decades, and is not just mechanically picking up returning (male) migrants. Large significant effects are seen for women, as well as for children under age 5 (perhaps suggestive of a small baby boom post-migration shock). Thirty years after the labor migration episode, total population in the high *K/L shock* districts had increased by around 5.2%.

Figure 1b shows the spatial reorganization of population within districts in the wake of the shock, where the comparison year is 1966. By 1977, districts that received 73 USD more per worker than other districts had 1.3 percentage points more of their population living in an urban area. This significant gap in the share of population living in urban areas persists through to 2008. Internal population rearrangements within districts are therefore likely to have contributed to the observed structural change in these rural labor markets.

6 Unpacking timing and persistence of accumulation

The results in Section 5 demonstrate that the *K/L shock* impacted the structure of work in rural labor markets, but this impact took time to be felt in any magnitude. Although population growth effects appear immediately after the shock, the shifts in employment across sectors take longer to manifest.

Section 1 outlined channels through which temporary migrant capital injections might trigger sectoral shifts in employment shares, and why district-level shifts may only be seen in the long-run. The main channels of impact included: (1) an initial positive demand shock increasing the consumption of farm and non-farm goods; (2) an increase in the supply of liquidity facilitating farm and non-farm investments; and (3) an increase in human capital investments. These three channels may not occur at all, may occur independently, or may be related: for example, the initial consumption boost may have an immediate impact on demand that becomes a persistent effect when coupled with the second and third channels. The third channel additionally complements the second channel in the future, when the next generation of workers enters the labor market.

In this section, we show that the pattern of long-run impacts we estimate in Section 5 are consistent with migrant capital operating through several of the mechanisms from Section 2. Throughout most of this section, we estimate regressions that take the form of Equation (1). For the outcomes considered in this section, we have the added benefit of using pre-shock data on outcomes so that we can estimate standard difference-in-differences models.

6.1 Evidence on increased demand for farm and non-farm capital

Ideally we would have data to show short-run purchases of food and non-farm goods increasing immediately following the *K/L shock*. Unfortunately, in the absence of annual data on any of these outcomes, we can only observe outcomes for non-food outcomes from 1987 onwards. Using these longer-run outcomes, we can test for persistent changes in holdings of farm and non-farm goods.

In Table 5 we examine whether there is a persistent increase in purchases of both farm and non-farm goods. Outcomes for Panel A, farm goods, are from the National Sample Survey of Agriculture in 1968 and the National Household Income and Expenditure Survey in 1998. We weight up to district-level using sample weights. Outcomes for Panel B, non-farm goods, are from the national Census in various years. In both panels, we estimate how farm and non-farm goods' ownership changes in districts exposed to high versus low *K/L shock*, in the decades after the plane crash, controlling for differences in ownership across these districts in the decades before the labor ban.

Panel A presents estimates of the effect of exposure to the migrant capital shock on farm assets. We measure the average number of hoes and pangas per household before, and twenty years after, the migration shock; and do the same for the share of households owning any cattle. Oxcart ownership can only be measured in 1987, 1998 and 2008.

Districts receiving more migrant capital per worker do not seem to be strongly investing in farm-specific capital over the longer run. Although the coefficients on *K/L shock* in 1987 are positive for hoes, pangas, and cattle, these estimates are not statistically significant using randomization inference p-values.[18] We do not place much weight on the differential decline in oxcart use in high *K/L shock* districts, since this effect is so small.

In contrast to these results on farm-specific capital, there is some evidence that non-farm specific assets increased, post-migration shock. In Table 5, Panel B, we examine changes in ownership of assets that could be used for household consumption or for non-farm work. We measure the share of households in the district in a given year that own a radio, have a household with durable walls, a durable roof, and/or both durable walls and a durable roof. Panel B shows immediate impacts on radio ownership in districts receiving more migrant capital per work. Ten years after the plane crash, we see significant increases in the share of households with durable walls and roofs, and durable roofs alone, in districts with higher *K/L shock*. The share of houses constructed with better quality materials increased by 4.8 percentage points (0.054*73), or about 34%, in the decade following migrant capital inflows.

There is little to suggest that the temporary increase in migrant capital per worker had a persistent impact on agricultural practices in rural Malawi; changes in agricultural productivity are unlikely to account for the sectoral employment shifts we measure in Section 5. Rather, the results in Table 5 indicate a meaningful change in housing quality that may well have resulted from an increase in the demand for non-farm goods. Improved housing represents increased consumption. But it also represents increased investment in premises required to provide non-farm goods and services (e.g. spaza shops, village bars and food

[18]Data on inputs like fertilizer, hybrid seeds, or type of crop planted where we may see a longer term impact are not available prior to the early 2000s.

stalls, hairdressing premises etc). Any positive demand shock triggered by the migrant capital inflows post-plane crash would likely have stimulated the non-farm goods and services sectors, and could have led to increased investment in these non-farm sectors.

Complementing this analysis of asset ownership in the Census data, we use the 1992 Malawi Demographic Health Survey (DHS) to capture a larger set of household assets and to measure these assets 15 years after the end of the migration shock. If our hypothesis about the migrant capital shock relaxing liquidity constraints is correct, then we should expect to see increases in ownership across a wide range of assets, some of which might help with production in a household business, while others might generate more consumption benefits.

The 1992 Malawi DHS is a cross-sectional dataset, so we estimate the cross-sectional correlation between measures of household assets in 1992 and the *K/L shock* received by each district in the mid-1970s. We ask whether districts that received more capital after the migration shock exhibit greater physical asset ownership across a wider range of assets. For households h in district d, we estimate:

$$HHAssets_{hd} = \gamma_0 + \gamma_1 \frac{K_d}{L_d} shock + W_d \sigma + \mu_r + \omega_{hd} \qquad (2)$$

where the W_d variables are as before. Since we cannot include district fixed effects, we rely on baseline district-level controls and region fixed effects (μ_r) to account for any differences in initial wealth conditions across districts. We also control for the number of migrants returning to each district prior to 1974 (in the 1967-1973 window), to compare districts with the same migration “propensity”. ω_{hd} an idiosyncratic error term. Our outcomes $HHAssets_{hd}$ include indicators for ownership of specific assets like electricity, radios and cars, as well as the DHS wealth index (units are in standard deviations) and a count of household-level assets. Randomization inference p-values are provided for each coefficient.

Table 6 column (1) shows that districts receiving more *K/L shock* by 1977 have higher asset holdings 15 years later, as measured by a higher asset index, more durable houses (roofs and floors), greater access to electricity, and higher radio and car ownership. These results indicate changes in household fortunes that occur after a shorter time period than the 20 years for which we see some of the main employment results in the Census data. This table helps us understand that some of the "delay" in employment results is a function of not having outcomes data for intervening Census years.[19]

The DHS data allow us to further investigate whether all of the gains from the migrant capital shock are limited to the older generation in 1992; those who would have been adults at the time of the shock. This investigation helps us further understand the ways in which the migrant capital shock persisted. In Table 6 columns (2) and (3), we split the sample into household heads that were adults at the start of the 1967 Treaty and those who would have been children, or not yet born, at the time of the migration shock.

Comparing columns (2) and (3) in Table 6 shows that both age groups enjoyed some higher asset ownership in high *K/L shock* districts. Specifically, 15 years after the shock, older household heads from higher *K/L shock* districts, have more assets, higher quality households with more services, and better transportation (more cars, fewer bikes) than older household heads in lower capital districts. This gives us direct evidence of physical capital accumulation that is independent of the next generation's potential human capital accumulation. At the same time, column (3) shows that younger household heads *also* enjoyed better quality houses and (some) better services (e.g. water access) in high migrant capital districts, although their relative asset accumulation is smaller than that of the older households. This table suggests that human capital accumulation among younger cohorts

[19]As Section 2 noted, labor productivity in Malawi is estimated to be 2–4 times higher in non-farm than in farm sectors. The wealth gains that we estimate at household level for a given district's mean exposure to migrant capital inflows therefore seem plausible, given these large potential average gains from switching.

may have played a role in shifting employment across sectors.

6.2 Evidence on human capital accumulation

We know from prior work (Dinkelman and Mariotti (2016)) that districts more exposed to the total migration shock 1967-1977 invested more in the schooling of those who were children at the time of the migration shocks.

In Table 7, we relate education profiles of different districts and cohorts to the inflow of post-plane crash migrant money per worker. We focus on total accumulated education of adults aged 20 to 65 in the 1998 census, some of whom would have been eligible for schooling during parts of the migration surge. We ask whether mean levels of human capital are higher for cohorts living in districts that received the largest *K/L shock* during their years of primary school eligibility.

We estimate the following regression for the average education outcomes of cohort c in district d ($Educ_{cd}$), separately by gender:

$$Educ_{cd} = \gamma_0 + \gamma_1 \frac{K_d}{L_d} * Shock_c + \gamma_2 \frac{K_d}{L_d} * Postshock_c + \phi_c + \mu_d + V_d Cohort_c \sigma + \omega_{cd} \quad (3)$$

We exploit an extra piece of time variation to check whether cohorts eligible for primary schooling in the narrow 1974-1977 period ($Shock_c$ cohorts) and right after the end of the shock, 1977-1980 ($Postshock_c$ cohorts), gained more schooling if they were in high capital inflow districts.[20] Individuals are grouped into nine different five-year age bins. Primary school age eligibility in Malawi runs from age 6 to age 15. Our control cohorts (omitted cohort category) are those eligible for primary schooling before 1974. Cohort dummies ($Cohort_c$) are interacted with control variables in V_d: the log of 1945 population density, the

[20]Effects on education found in the post-shock cohort would lend support to our claim that the migrant capital shock did have immediate impacts, but that we just can't measure them in any other data other than in long-run education outcomes.

1945 literate youth share, and region fixed effects. μ_d are district fixed effects.

Results in Table 7 show that for each additional 73 USD received by returning migrants in a district, female education rose by a significant 0.15-0.16 years for the shock-era and post-shock era cohorts (0.73*0.23 and 0.73*0.21). Male education rose by a smaller and non-significant 0.05-0.09 in the same cohorts. Each of these effects is relative to the control cohort of older adults ineligible for schooling at the time of the labor ban.[21] Migrant capital raises the average female level of education by 8 - 8.5%, and for makes by a smaller 0.4 - 2.7%. Exposure to more migrant capital also raises female enrollment in primary school by 3.5 - 4% (significant for the post-shock cohorts), with negligible changes in enrollment for males. And, the gender differences in the impact on education are statistically significant.

The results of Table 7 are important for several reasons. Together with the results in Tables 5 and 6, they indicate that the migrant capital coming back to Malawi in the wake of the plane crash was not just consumed, but also invested. In addition, the individuals who gained the most education in response to more *K/L shock* (younger females) would only have entered the work force in early adulthood around the mid- to late 1990s. This is exactly the decade in which we start to see larger, significant impacts on sectoral employment shifts, and more so for women than men. It is likely that the human capital investments triggered by migrant capital contributed to the structural shifts of labor out of agriculture and towards services. This is in line with other micro-economic evidence that suggests that people with more schooling are more likely to work off-farm relative to on-farm (e.g. Huffman (1980) for the historical US, Yang (1997) for China and Fafchamps and Quisumbing (1999) for Pakistan).

[21]These estimates likely underestimate the impact of total migrant capital per worker on total educational attainment at district level, because some of the control cohorts would have been eligible for schooling in the 1967-1973 window.

6.3 Discussion

The three sets of results presented in this section support our accumulation hypothesis and explain why the sectoral employment shifts take time to gain traction at the level of the district, and why the initial impact of the remitted migrant capital is increasing over time.

First there are small direct impacts on the consumption of non-farm goods (radios, household construction). This is consistent with our claim that under non-homothetic preferences, more income will translate into increased demand for non-food items. This increase in demand may cause some of the initial shift we see in employment shares.

We claimed in Section 1 that for this shift in demand to be permanent we need workers in the market positioned to take advantage of the demand shock through access to capital that they either lend out or borrow to invest in their businesses. We see evidence of an increase in physical asset ownership that may be consistent with this second channel fostering accumulation. Twenty years after the shock, the employment share effect has grown as a result of both the initial trigger to demand as well as any subsequent increased investment.

While these initial liquidity constraints were being resolved, workers were also investing in the human capital of the next generation. This generation would have joined the labor market anywhere between 10 and 25 years after the shock and contributed to the growth in the effect sizes we see two and three decades after the shock.

We do not distinguish between the share of employment reallocating across sectors that is caused by each channel. Rather we note that all three together provide evidence consistent with the conceptual channels in Section 1. More importantly, all three are consistent with the speed at which we see the employment share shifts taking place.

A final question is why we see employment shifts out of agriculture and into services that are larger for women than men. Although our data do not have sufficient variation to always distinguish statistically between male and female impacts, the point estimates for female employment shares are larger than for male employment shares in almost all cases. We note that differences between men's and women's work through the structural transformation of an economy are not well understood in African settings. This makes it difficult to know whether we should expect larger responses to the labor migration shock from men or from women. While women tend to be marginal or added workers in agricultural models of household production (e.g. Schultz (2001), working less in non-farm family businesses, it is not clear how we should expect female employment to change at a more aggregate level, as economies grow richer. Reviewing the literature on structural transformation reveals a significant gap in the literature on this front.

Although we have come to expect a U-shaped pattern of female labor force participation that is seen through the structural transformation (Goldin, 1995), empirically, these patterns are mainly observed in historically rich countries, in which manufacturing has been an important sector of work through the structural transformation. As we noted in section 2, Malawi's structural transformation resembles patterns in the rest of Africa (McMillan, Rodrik and Verduzco-Gallo, 2014), where manufacturing has not been as important a sector of work. If women have comparative advantage in services (e.g. Ngai and Petrongolo (2017) and Ngai, Olivetti, and Petrongolo (2021)), and if the way that African economies develop is to skip over manufacturing, then it does not seem unreasonable that female employment shifts out of agriculture and into low-skilled services in response to the labor migration shock in our Malawi setting. We leave exploration of this topic to future work.

7 Conclusion

This paper marshals historical data from Malawi to exploit a natural experiment that shocked access to international labor migration to demonstrate that migrant capital can contribute to long run changes in the structure of rural labor markets. In places receiving larger inflows of migrant capital per worker, employment shifts out of agriculture and towards the service sector. This is particularly the case for women. Jobs in construction, retail, general services, and transport and communications increased and employment became more diverse in those districts that received more capital per worker after the migration shock. Even after the end of migration, accumulation persisted at higher rates. Districts with more migrant capital own more physical, non-farm capital and have higher human capital over the long run.

Our work sheds light on a relatively understudied period in Malawi's economic history and is broadly relevant to African labor markets in the past and present. Many southern African countries were affected by similar fluctuations in worker flows to the South African gold mines. Structural change could also have occurred in these other countries as a result of capital accumulated through international labor migration.

Although historical, our work is relevant for countries considering temporary or seasonal labor migration programs. Legal, time-limited migration might present a practical way for communities to accumulate capital in labor-rich, resource-poor countries, with important implications for women's work in these countries. When such migration flows are widespread, and accompanied by large return flows of money, impacts on the local labor market can be persistent, with positive long-term consequences for labor allocation across sectors at home.

References

Alvarez-Cuadrado, Francisco and Francesco Amodio and Markus Poschke, 2021. "Selection and Absolute Advantage in Farming and Entrepreneurship", Working Paper.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak and Anton Heil, 2022. "Why Do People Stay Poor?", *The Quarterly Journal of Economics* 137 (2): 785-844.

Beegle, Kathleen, Emanuela Galasso and Jessica Goldberg. 2017. "Direct and Indirect Effects of Malawi's Public Works Program on Food Security", *Journal of Development Economics* 128: 1-23.

Buera, Francisco J., Joseph P. Kaboski and Yongseok Shin. 2014. "Macro-perspective on asset grants programs: Occupational and wealth mobility", *American Economic Review Papers and Proceedings* 104 (5): 159–164

Bustos, Paula, Bruno Caprettini and Jacopo Ponticelli. 2016. "Agricultural Productivity and Structural Transformation. Evidence from Brazil", *American Economic Review*, 106(6): 1320-1365

Bustos, Paula, Gabriel Garber, and Jacopo Ponticelli. 2020. "Capital Accumulation and Structural Transformation", *Quarterly Journal of Economics*, 135 (2): 1037–1094.

Chen, Chaoran and Diego Restuccia and Raul Santaeulalia-Llopis. 2023. "Land Misallocation and Productivity", *American Economic Journal: Macroeconomics*, 15(2): 441-465.

Clemens, Michael and David McKenzie. 2018. "Why Don't Remittances Appear to Affect Growth?", *The Economic Journal* 128(612) *https://doi.org/10.1111/ecoj.12463*

Davidson, R. and James G. MacKinnon. 2010. "Wild Bootstrap Tests for IV Regression.", *Journal of Business and Economic Statistics* 28: 128–144.

Dinkelman, Taryn and Martine Mariotti. 2016. "The Long-Run Effects of Labor Migration on Human Capital Formation in Communities of Origin", *American Economic Journal: Applied Economics*, 8 (4): 1-35.

Dustmann, Christian, and Oliver Kirchkamp. 2002. "The Optimal Migration Duration and Activity Choice after Re-migration", *Journal of Development Economics* 67 (2): 351-72.

Erten, Bilge and Jessica Leight. 2021. "Exporting Out of Agriculture: The Impact of WTO Accession on Structural Transformation in China", *Review of Economics and Statistics* 103 (2): 364–380.

Fafchamps, Marcel and Agnes Quisumbing. 1999. "Human Capital, Productivity, and Labor Allocation in Rural Pakistan", *Journal of Human Resources*, 34(2)

Foster, Andrew D. and Mark R. Rosenzweig. 2008. "Economic Development and the Decline of Agricultural Employment" in Handbook of Development Economics Vol.4: Chapter 47. Elsevier.

Gibson, John, David McKenzie and Halahingano Rohorua. 2014. "Development Impacts of Seasonal and Temporary Migration: A Review of Evidence from the Pacific and Southeast Asia", *Asia and the Pacific Policy Studies* 1(1): 18-32

Gibson, John and David McKenzie. 2014. "The Development Impact of a Best Practice Seasonal Worker Policy", *Review of Economics and Statistics*, 96(2): 229-243

Goldberg, Jessica. 2016. "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi", *American Economic Journal: Applied Economics* 8: 1

Goldin, Claudia. 1995. "The U-Shaped Female Labor Force Function in Economic Development and Economic History" in Schultz T.P. 1995. Investment in Women's Human Capital and Economic Development. University of Chicago Press. pp. 61-90

Gollin, Douglas, Stephen L. Parente, and Richard Rogerson. 2002. "The Role of Agriculture in Development" *American Economic Review Papers and Proceedings* 92: 2

Gollin, Douglas, David Lagakos and Michael E. Waugh. 2014. "The Agricultural Productivity Gap", *Quarterly Journal of Economics*, 129(2): 939-993

Gollin, Douglas, Remi Jedwab and Dietrich Vollrath. 2016 "Urbanization With and Without Industrialization", *Journal of Economic Growth* 21 (1): 35-70

Herrendorf, Berthold, Richard Rogerson and Àkos Valentinyi. 2014. "Growth and Structural Transformation", in (eds.) Philippe Aghion and Steven N. Durlauf, Handbook of Economic Growth Volume 2, Elsevier.

Huffman, William. 1980. "Farm and Off-Farm Work Decisions: The Role of Human Capital", *Review of Economics and Statistics*, 62(1): 14-34

Kaboski, Joseph P. and Robert M. Townsend. 2012. "The Impact of Credit on Village Economies", *American Economic Journal: Applied Economics* 4(2), pages 98-133, April.

Khanna, Gaurav, Emir Murathanoglu, Caroline Theoharides, and Dean Yang 2022. "Abundance from Abroad: Migrant Income and Long-Run Economic Development", (NBER Working Paper No. 29862).

Kinnan, Cynthia, Shing-Yi Wang and Yongxiang Wang. 2018. "Access to Migration for Rural Households", *American Economic Journal: Applied Economics*, 10 (4): 79–119.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labor", *The Manchester School*, 22(2): 139-91

Mariotti, Martine. 2015. "Fathers' Employment and Sons' Stature: The Long-Run Effects of a Positive Regional Employment Shock in South Africa's Mining Industry", *Economic Development and Cultural Change*, 63(3): 485-514.

McCaig, Brian and Nina Pavcnik. 2018. "Export Markets and Labor Allocation in a Low-Income Country", *American Economic Review* 108(7): 1899-1941.

McMillan Margaret, Dani Rodrik and *Íñigo* Verduzco-Gallo. 2014. "Globalization, Structural Change, and Productivity Growth, with an Update on Africa", *World Development* 63: 11–32.

Ngai, L. Rachel, Claudia Olivetti and Barbara Petrongolo. 2021. "Structural transformation and U-shaped female employment", Working Paper.

Ngai, L. Rachel and Barbara Petrongolo. 2017. "Gender Gaps and the Rise of the Service Economy", *American Economic Journal: Macroeconomics* 9 (4): 1-44.

Porzio Tommaso and Federico Rossi and Gabriella Santangelo. 2022. "The Human Side of Structural Transformation",*American Economic Review* 112 (8): 2774-2814.

Robinson, Amanda Lea. 2016. "Internal Borders: Ethnic Diversity and Market Segmentation in Malawi", *World Development* 87:371–384.

Roodman, David, James G. MacKinnon, Morten Ørregaard Nielsen, and Matthew D. Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata Using Boottest", *The Stata Journal* 19 (1): 4—60 10.1177/1536867X19830877

Rosenstein-Rodan, P. N. 1943. "Problems of Industrialisation of Eastern and South-Eastern Europe"', *Economic Journal*, 53 (202): 202-211

Schultz, T. Paul. 2001 "Women's Roles in the Agricultural Household: Bargaining and Human Capital Investments," in Handbook of Agricultural Economics, Vol. 1, ed. by B. Gardner and G. Rausser. Amsterdam: North-Holland.

Theoharides, Caroline. 2018. "Manila to Malaysia, Quezon to Qatar: International Migration and its Effects on Origin-Country Human Capital", *Journal of Human Resources*, Vol. 53. No. 4, Fall 2018, pp. 1022-1049.

Theoharides, Caroline. 2020. "The Unintended Consequences of Migration Policy on Origin-Country Labor Market Decisions", *Journal of Development Economics*, 142

Treaty Series No.10/1967. 1967. "Agreement between the Governments of the Republic of South Africa and Malawi relating to the Employment and Documentation of Malawi Nationals in South Africa." Place and date of signing: Pretoria and Blantyre, 1 August 1967, Date of entry in force: 1 August 1967.

Woodruff, Christopher, and Rene Zenteno. 2007. "Migrant Networks and Microenterprises in Mexico.", *Journal of Development Economics*, 82 (2): 509-28.

World Bank. 2014. "Global Financial Development Database (GFDD)", downloaded from *http://data.worldbank.org/data-catalog/global-financial-development* on August 9, 2016

World Bank. 2016. "World Bank National Accounts data, and OECD National Accounts data files", data downloaded from *http://data.worldbank.org/indicator/NV.SRV.TETC.ZS?locations=MW* on September 2, 2016

World Bank. 2019. "Migration and Remittances: Recent Developments and Outlook", *Migration and Development Brief Num. 31*, April

Yang, Dennis. 1997. "Education and Off-Farm Work", *Economic Development and Cultural Change* 45 (3): 613-632.

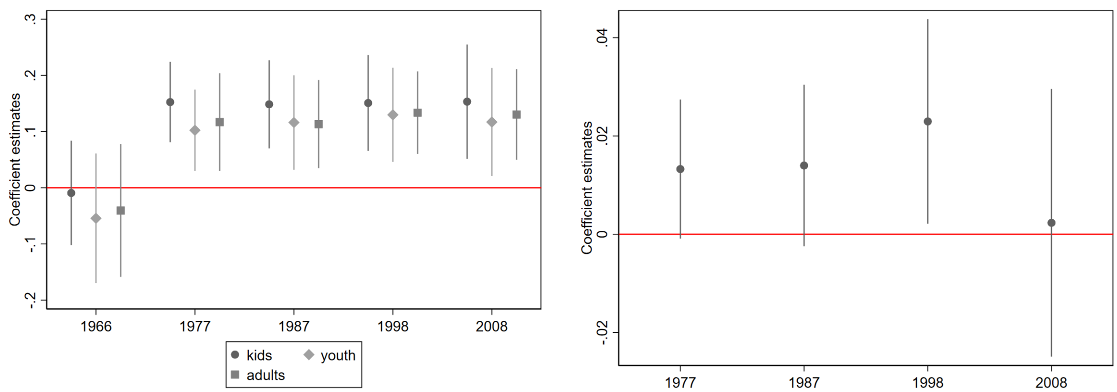
Yang, Dean. 2008. "International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks", *Economic Journal* 118 (528): 591–630.

Young, Alwyn. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results", *Quarterly Journal of Economics* 134 (2):557–598.

RESULTS APPENDIX: FIGURES AND TABLES

Figure 1: Effects of K/L Shock on District-level Population Growth and Urbanization

(a) Cohort-specific population density **(b) Urbanization**



Figures plot the coefficients from Equation (1), for equations where the outcomes are the log of the age-specific population totals in the district (1a) or the share of the urban population in the district (1b). Age groups are: adults (19+), youth ($5-18$) and children (under age 5). Points shown are coefficients on the interaction of Census year dummies with the district-level *K/L shock* variable. Base year is 1945. Mean value of *K/L shock* is 73 USD. Full regression results are in Online Results Appendix Table A.3.

Table 1: Correlates of District-Level K/L Shock and K/L Flows

	K/L Shock	**K/L Total**
Log Pop Density 1945	-0.816	-1.296
	[0.132]	[0.158]
Central Region [0/1]	0.478	0.919
	[0.242]	[0.226]
Southern Region [0/1]	0.299	0.340
	[0.400]	[0.446]
Married 1977 [0/1]	-5.527	-8.838
	[0.278]	[0.290]
Youth Literacy Rate 1945	-9.123	-16.248
	[0.236]	[0.216]
Malaria area [0/1]	-1.911	-3.052
	[0.078]	[0.110]
Estate District [0/1]	-0.624	-1.080
	[0.186]	[0.184]
Male Share No Wage 1966	-17.978	-27.660
	[0.108]	[0.138]
Female Share No Wage 1966	11.293	17.712
	[0.178]	[0.194]
Maize Suitability	-0.006	-0.017
	[0.446]	[0.426]
Tobacco Suitability	-0.002	0.003
	[0.530]	[0.460]
Observations	24	24
Y Mean	0.730	1.294
R2	0.586	0.551
Joint Test	0.650	0.760

Each column shows coefficients from a regression of the migrant capital per worker outcome on the full set of geographic and baseline district-level covariates. Randomization inference p-values are shown in square brackets, and for the joint test. Results are district population weighted (1977 weights). Unit of observation is the district. *K/L* variables are scaled by 100USD per worker.

Table 2: Long-run impacts of *K/L shock* on sectoral employment shares

	Agric.	Manuf.	Services	Ind. Index
Panel A: Share of Women in Each Industry				
(K/L)Shock* Three Decades Post	-.044	.0053	.049	-.0077
	[0.025]	[0.080]	[0.013]	[0.011]
(K/L)Shock* Two Decades Post	-.037	.0022	.025	-.036
	[0.015]	[0.030]	[0.013]	[0.037]
(K/L)Shock* One Decade Post	-.009	.0022	.0052	-.014
	[0.005]	[0.384]	[0.004]	[0.245]
N obs	96	96	96	96
Y Mean	0.878	0.019	0.076	0.796
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.104	0.332	0.104	0.162
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.020	0.148	0.036	0.162
Panel B: Share of Men in Each Industry				
(K/L)Shock* Three Decades Post	.0022	.0022	.0063	.028
	[0.047]	[0.002]	[0.003]	[0.301]
(K/L)Shock* Two Decades Post	-.019	.0048	.018	-.011
	[0.018]	[0.016]	[0.002]	[0.043]
(K/L)Shock* One Decade Post	-.011	.0011	.0038	-.014
	[0.004]	[0.025]	[0.001]	[0.218]
N obs	96	96	96	96
Y Mean	0.728	0.087	0.152	0.577
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.054	0.758	0.136	0.044
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.446	0.752	0.236	0.924
Panel C: Testing female vs male (r.i. p-values)				
$\alpha_{3,F} = \alpha_{3,M}$	0.984	0.994	0.608	0.034
$\alpha_{2,F} = \alpha_{2,M}$	0.986	0.020	0.978	0.974
$\alpha_{1,F} = \alpha_{1,M}$	0.002	0.988	0.980	0.002

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from May 1974 to Nov. 1975/N. miners returning over same time. Mean *K/L shock* = 0.73, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, interactions of linear trend with all baseline district variables in Table 1. Ind. Index is a Herfindahl index measuring sectoral concentration of work.

Table 3: Long-run impacts of *K/L* on sectoral employment shares

	Agric.	Manuf.	Services	Ind. Index
Panel A: Share of Women in Each Industry				
(K/L)* Three Decades Post	-.025	.0031	.028	-.0043
	[0.032]	[0.161]	[0.033]	[0.019]
(K/L)* Two Decades Post	-.021	.0013	.014	-.021
	[0.024]	[0.036]	[0.031]	[0.045]
(K/L)* One Decade Post	-.0053	.0014	.0028	-.0087
	[0.010]	[0.503]	[0.016]	[0.278]
N obs.	96	96	96	96
Y Mean	0.878	0.019	0.076	0.796
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.120	0.406	0.126	0.250
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.030	0.142	0.042	0.178
Panel B: Share of Men in Each Industry				
(K/L)* Three Decades Post	.00073	.0016	.0033	.015
	[0.052]	[0.004]	[0.004]	[0.336]
(K/L)* Two Decades Post	-.012	.0031	.011	-.008
	[0.024]	[0.025]	[0.003]	[0.042]
(K/L)* One Decade Post	-.0075	.00095	.0022	-.0097
	[0.007]	[0.087]	[0.003]	[0.236]
N obs.	96	96	96	96
Y Mean	0.728	0.087	0.152	0.577
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.036	0.736	0.116	0.024
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.374	0.702	0.250	0.904
Panel C: Testing female vs male (r.i. p-values)				
$\alpha_{3,F} = \alpha_{3,M}$	0.984	0.994	0.554	0.040
$\alpha_{2,F} = \alpha_{2,M}$	0.984	0.020	0.974	0.972
$\alpha_{1,F} = \alpha_{1,M}$	0.000	0.994	0.976	0.000

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from Aug. 1967 to Nov. 1975/N. miners returning over same time. Mean *K/L shock* = 1.29, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, interactions of linear trend with all baseline district variables in Table 1. Ind. Index is a Herfindahl index measuring sectoral concentration of work.

Table 4: Long-run impacts of *K/L shock* on non-farm sectors

	General Manuf.	Const'n	Gen. Services	Retail	Trans.and Comms
Panel A: Share of Women in Each Industry					
(K/L)Shock* Three Decades Post	.0031	.002	.014	.034	.00059
	[0.018]	[0.081]	[0.029]	[0.080]	[0.051]
(K/L)Shock* Two Decades Post	.0012	.00087	.0067	.018	.00045
	[0.010]	[0.029]	[0.010]	[0.150]	[0.052]
(K/L)Shock* One Decade Post	.0014	.00075	.0026	.002	.00054
	[0.001]	[0.061]	[0.003]	[0.291]	[0.054]
N. obs	96	96	96	96	96
Y Mean	0.014	0.005	0.029	0.046	0.001
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.588	0.654	0.270	0.038	0.008
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.278	0.468	0.106	0.030	0.032
Panel B: Share of Men in Each Industry					
(K/L)Shock* Three Decades Post	-.0039	.0073	-.0072	.018	-.0047
	[0.047]	[0.167]	[0.008]	[0.071]	[0.047]
(K/L)Shock* Two Decades Post	-.0077	.013	.0019	.022	-.0053
	[0.031]	[0.006]	[0.000]	[0.283]	[0.025]
(K/L)Shock* One Decade Post	-.0039	.0049	-.00042	.0053	-.0011
	[0.005]	[0.083]	[0.001]	[0.477]	[0.027]
N. obs	96	96	96	96	96
Y Mean	0.049	0.036	0.074	0.065	0.013
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.290	0.040	0.156	0.038	0.008
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.428	0.150	0.504	0.088	0.064
Panel C: Testing female vs male (r.i. p-values)					
$\alpha_{3,F} = \alpha_{3,M}$	0.882	0.018	0.004	0.980	0.024
$\alpha_{2,F} = \alpha_{2,M}$	0.058	0.006	0.278	0.012	0.004
$\alpha_{1,F} = \alpha_{1,M}$	0.138	0.010	0.032	0.016	0.926

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from May 1974 to Nov. 1975/N. miners returning over same time. Mean *K/L shock* = 0.73, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, interactions of linear trend with all baseline district variables in Table 1. Ind. Index is a Herfindahl index measuring sectoral concentration of work.

Table 5: Long-run impacts of *K/L shocks* on farm and non-farm physical capital

	Farm Capital			
	Pangas(*)	**Hoes(*)**	**Any Cattle**	**Oxcart**
Panel A: Share of households with (number of*)				
(K/L)Shock* Three Decades Post				-.0053
				[0.056]
(K/L)Shock* Two Decades Post	.024	.061	.016	-.0027
	[0.220]	[0.224]	[0.258]	[0.165]
Y-mean	0.282	1.463	0.094	0.015
Base Year	1968	1968	1968	1987
Years of Data in Sample	68, 77	68, 77	68, 77	87, 98, 08
N. obs	46	46	46	69
Joint Test: (K/L) Shock* Decade	n/a	n/a	n/a	0.000

	Non-Farm Capital			
	Radio	**Durable Walls**	**Durable Roof**	**Durable Roof and Walls**
Panel B: Share of Households With:				
(K/L)Shock* Three Decades Post	-.00043	.024	.043	.048
	[0.489]	[0.285]	[0.091]	[0.101]
(K/L)Shock* Two Decades Post	.0084			
	[0.193]			
(K/L)Shock* One Decade Post	.0081	.067	.053	.054
	[0.122]	[0.068]	[0.018]	[0.022]
(K/L)Shock*Post	.0067			
	[0.046]			
Y-mean	0.277	0.389	0.133	0.124
Base Year	1968	1968	1968	1968
Years of Data in Sample	1969-2008	69, 87, 08	69, 87, 08	69, 87, 08
N. obs	115	69	69	69
Joint Test: (K/L) Shock* Decade	0.000	0.000	0.000	0.000

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from May 1974 to Nov. 1975/N. miners returning over same time. Mean *K/L shock* = 0.73, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, interactions of linear trend with all baseline district variables in Table 1. Ind. Index is a Herfindahl index measuring sectoral concentration of work.

Table 6: Impacts of *K/L shock* on wealth 15 years later: Evidence from Malawi's 1992 DHS

	Full Sample N=5,323	Oldest HHs: Head $\geq$ 38 y.o N = 3,035	Youngest HHs: Head $<$ 38 y.o N = 2,288
DHS Wealth Index (s.d)	.12	.14	.093
	[0.654]	[0.667]	[0.700]
Count of assets	.21	.23	.17
	[0.340]	[0.300]	[0.300]
Durable Roof	.08	.082	.074
	[0.246]	[0.100]	[0.200]
Durable Floor	.061	.066	.053
	[0.158]	[0.067]	[0.067]
Electricity	.0093	.016	-.0018
	[0.072]	[0.067]	[0.067]
Radio	.0047	.015	-.0033
	[0.060]	[0.067]	[0.033]
Car	.005	.0095	.00034
	[0.022]	[0.033]	[0.033]
Motorbike	-.002	-.0019	-.0022
	[0.046]	[0.100]	[0.033]
Bike	-.051	-.048	-.05
	[0.004]	[0.000]	[0.033]
Improved Toilet	-.013	-.019	-.0055
	[0.002]	[0.000]	[0.033]
Improved Water Source	.042	.038	.042
	[0.010]	[0.000]	[0.000]

Table shows coefficients from separate (DHS-weighted) regressions of each asset variable on the *K/L shock* variable measured in units of 100USD per worker (Equation (2)). DHS wealth index is a composite index of assets constructed using principal components analysis. Durable roof through Improved Water Source outcomes are binary. Randomization inference p-values are in square brackets. All regressions control for region fixed effects and baseline district-level variables: total number of migrants 1967-1977, population density in 1945, literacy rates, an indicator for estates in the district, a malaria risk indicator, the share of men and women married in 1977 and the share of men and women working for no cash wage in 1966.

Table 7: Long-Run Impacts of *K/L Shock* on Investment in Human Capital of the Next Generation, by Gender

	Yrs of Completed Education		Any Primary Schooling	
	Females	Males	Females	Males
(K/L) Shock *	.23	.13	.019	.0022
Shock Cohorts, γ_1	[0.064]	[0.418]	[0.182]	[0.932]
(K/L) Shock *	.21	.068	.017	-.0047
Post-Shock Cohorts, γ_2	[0.000]	[0.658]	[0.002]	[0.852]
N. obs	240	240	240	240
Y mean	1.881	3.231	0.347	0.474
Joint test:	0.046	0.596	0.044	0.686
Testing female vs male (r.i. p-values)				
$\gamma_{1,F} = \gamma_{1,M}$	0.000		0.000	
$\gamma_{2,F} = \gamma_{2,M}$	0.000		0.000	

The table presents coefficients from estimating Equation (3) in the text. Randomization inference p-values for each coefficient are presented in square brackets and all joint tests use r.i. p-values. *K/L shock* is deferred pay returning to each district between May 1974 and November 1975 divided by the number of miners returning in those years. Mean *K/L shock* = 0.73, units are 100USD per miner. Outcomes are constructed from Census 1977 and 1998. Unit of observation is the district-gender-cohort cell. Shock cohorts are age-eligible for primary school during 1974-1977. Post-shock cohorts are age-eligible for primary school 1977-1980. All regressions contain district fixed effects and cohort dummies. Other controls include a trend term interacted with baseline district variables: adult literacy in 1945, population density in 1945, and region dummies. Regressions are unweighted.

ONLINE RESULTS APPENDIX: NOT FOR PUBLICATION

This online appendix contains supplementary material for Dinkelman, Kumchulesi, and Mariotti. *"Labor migration, capital accumulation, and the structure of rural labor markets"*.

Figure A.1: Capital Flows and Migrant Workers Over Time

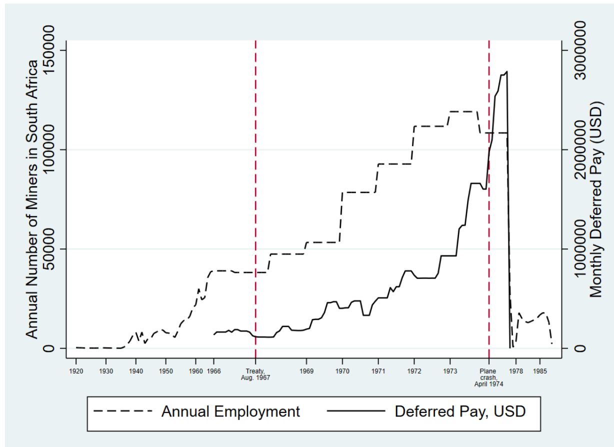
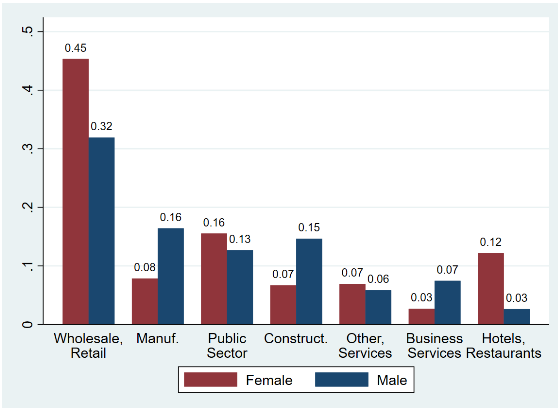
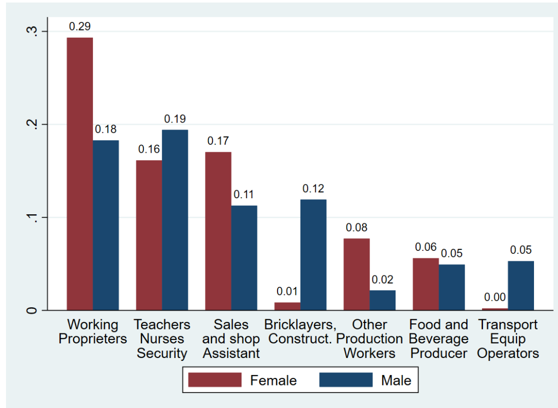


Figure shows annual number of Malawians contracted to work on SA mines and monthly deferred pay returned to Malawi. Vertical lines indicate (left to right) the treaty month (Aug. 1967) and the plane crash month (April 1974).

Figure A.2: Industries and Occupations among Non-farm Workers by Gender, 2008



(a) Industry of Work



(b) Occupation

Figure shows share of men and women employed in non-farm industries (2a) or non-farm occupations (2b) using two digit industry and occupation classifications in the 2008 Malawi Census.

Figure A.3: Annual Wages for African Mineworkers and Global Gold Prices

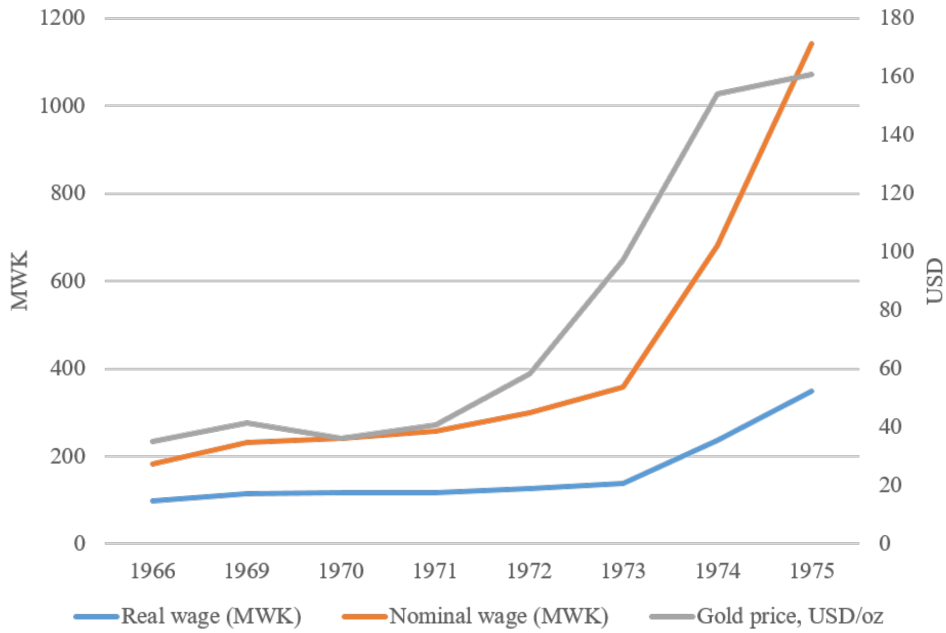


Figure plots annual average wages for mineworkers from Wilson (1972) for the years 1966 and 1969, and from Crush et al (1991) from 1970 onwards. Gold prices per ounce are from the National Mining Association, retrieved 17 March 2022: `https://nma.org/wpcontent/uploads/2016/09/historic_gold_prices_1833_pres.pdf`

Table A.1: **Employment Shares by Sector and Gender Over Time**

	1977	1987	1998	2008
Sector of work (Industry): Women				
Agriculture	0.943	0.949	0.888	0.708
Manufacturing	0.016	0.012	0.012	0.037
Services	0.028	0.032	0.067	0.198
Industrial Concentration Index	0.893	0.905	0.805	0.550
Sector of work (Industry): Men				
Agriculture	0.760	0.764	0.731	0.532
Manufacturing	0.093	0.076	0.074	0.132
Services	0.120	0.134	0.171	0.279
Industrial Concentration Index	0.618	0.625	0.589	0.357

Population-weighted shares of adults in each sector of work and employment category from Census data. Industry for the economically active population (workers and unemployed) 10 years and older is collapsed to 48 district-gender cells. Home workers are excluded from these definitions. Industry Concentration Index is a Herfindahl index of sector of work; larger values imply more concentration of work sector in the district. Online Data Appendix contains details of dataset construction. Totals do not sum to 1 because of residual 'not stated' categories for industry of work.

Table A.2: Summary statistics, district-level data

	Mean	SD	Min	Max
Components of migration shock				
N. men (aged 15-64), 1977	51,325.50	36,937.83	10,559.00	167,531.00
N. males ever abroad by 1977	19,556.63	15,421.10	4,232.00	75,324.00
Δ N. migrants, 1966-1977	13,641.73	10,666.94	2,815.97	50,120.54
Share male who were migrants, 1974-1977	0.06	0.01	0.03	0.09
Tot. deferred pay/district 1966-1975,USD mill.	2.25	3.53	0.00	16.29
Tot. deferred pay/person in district 1966-1975,USD	24.04	55.40	0.00	275.68
K/L Shock: capital/miner post-plane crash	0.73	1.10	0.00	5.63
K/L 1967-1974: capital/miner pre-plane crash	1.70	2.29	0.00	11.62
District-level descriptives at baseline				
Northern Region	0.21	0.41	0.00	1.00
Central Region	0.38	0.49	0.00	1.00
Southern Region	0.42	0.50	0.00	1.00
Pop. 1945	71,262.01	60,353.29	5,918.64	230,891.00
Pop. density per km^2, 1945	30.61	26.61	5.10	109.05
Share married in 1977	0.64	0.05	0.56	0.70
Share youth literate in English and vernacular, 1945	0.08	0.04	0.03	0.14
Share in high malaria area	0.28	0.35	0.00	1.00
Share with any agricultural estate	0.46	0.51	0.00	1.00
Share of men, no cash income in 1966	0.37	0.10	0.22	0.59
Share of women, no cash income in 1966	0.48	0.14	0.28	0.72
Mean Maize Suitability Index	36.08	14.23	8.84	71.55
Mean Tobacco Suitability Index	34.23	14.89	8.75	71.98

Raw means (unweighted) computed over 24 districts. Components of the migration shock variables are collected from administrative records and Census 1977. Baseline descriptives come from 1945, 1966, and 1977 Census data and from various geographic files for Malawi. Maize and tobacco suitability indices are computed from FAOSTAT data. Agricultural estate is a dummy variable indicating whether a district contains any cash crop estates (e.g. for tobacco or sugar).

Table A.3: Impacts of K/L Shock on Population Growth and Urbanization

	Ln pop	Ln Pop (Males Only)	Ln Pop (Females Only)	Ln Pop Age < 5	Ln Pop Age 5- 18	Ln Pop Age 18+	Share of Urban Pop
(K/L)Shock* Three Decades Post	.14	.16	.11	.16	.12	.13	.0023
	[0.069]	[0.246]	[0.195]	[0.067]	[0.098]	[0.098]	[0.055]
(K/L)Shock* Two Decades Post	.14	.17	.11	.16	.13	.14	.023
	[0.064]	[0.154]	[0.102]	[0.059]	[0.057]	[0.058]	[0.049]
(K/L)Shock* One Decade Post	.12	.15	.096	.15	.12	.12	.014
	[0.121]	[0.132]	[0.066]	[0.122]	[0.071]	[0.070]	[0.036]
(K/L)Shock* End of Migration	.12	.15	.087	.15	.1	.12	.013
	[0.126]	[0.099]	[0.057]	[0.128]	[0.068]	[0.068]	[0.055]
(K/L)Shock* One Decade Before	-.037	-.00056	-.074	-.0082	-.054	-.04	
	[0.382]	[0.338]	[0.352]	[0.385]	[0.400]	[0.380]	
Joint Test: K/L Shock*Decade, excluding the one decade before interaction							
All $\alpha's = 0$	0.120	0.206	0.206	0.126	0.578	0.608	0.044
Sum of $\alpha's = 0$	0.198	0.268	0.168	0.194	0.140	0.134	0.050

Randomization inference p-values for each coefficient are in square brackets and r.i. p-values are provided for all joint tests. K/L shock is the deferred pay returning to each district between May 1974 and November 1975 in millions of USD divided by the number of miners returning to Malawi in those years. The variable is measured in units of 100USD per miner. Mean of K/L shock is 0.73. Outcomes are measured in Census 1945, 1966, 1977, 1987, 1998, 2008 for all population density outcomes. Urban share is not available for 1945. Unit of observation is the district-year cell. All regressions include year and district fixed effects, and a full set of baseline controls from Table 1, interacted with a trend term.

Table A.4: Robustness: Long-run impacts of K/L shock on sectoral employment shares, Controlling for 2008 Voucher*2008 Decade

	Agriculture	Manufacturing	Services	Ind. Index
Panel A: Share of Women in Each Industry				
(K/L)Shock* Three Decades Post	-.045	.0065	.049	-.0098
	[0.021]	[0.081]	[0.011]	[0.014]
(K/L)Shock* Two Decades Post	-.038	.0032	.025	-.038
	[0.014]	[0.030]	[0.010]	[0.058]
(K/L)Shock* One Decade Post	-.0097	.0027	.0054	-.015
	[0.005]	[0.377]	[0.002]	[0.246]
N. obs	96	96	96	96
Y Mean	0.878	0.019	0.076	0.796
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.094	0.270	0.070	0.146
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.024	0.066	0.042	0.140
Panel B: Share of Men in Each Industry				
(K/L)Shock* Three Decades Post	-.0011	.0068	.0049	.024
	[0.043]	[0.002]	[0.003]	[0.323]
(K/L)Shock* Two Decades Post	-.022	.0087	.017	-.014
	[0.018]	[0.016]	[0.001]	[0.064]
(K/L)Shock* One Decade Post	-.013	.003	.0032	-.015
	[0.004]	[0.026]	[0.001]	[0.249]
N. obs	96	96	96	96
Y Mean	0.728	0.087	0.152	0.577
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.056	0.630	0.168	0.044
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.378	0.474	0.316	0.882
Panel C: Testing female vs male				
$\alpha_{3,F} = \alpha_{3,M}$	0.058	0.132	0.066	0.034
$\alpha_{2,F} = \alpha_{2,M}$	0.060	0.068	0.054	0.044
$\alpha_{1,F} = \alpha_{1,M}$	0.444	0.420	0.516	0.438

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from May 1974 to Nov. 1975/N. miners returning over same time. Mean *K/L shock* = 0.73, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, interactions of linear trend with all baseline district controls. **Regressions also control for 2008 fertilizer subsidy penetration interacted with a 2008 decade dummy.**

Table A.5: Robustness: Long-run impacts of K/L shock on sectoral employment shares, Non-linear controls*Trend

	Agriculture	Manufacturing	Services	Ind. Index
Panel A: Share of Women in Each Industry				
(K/L)Shock* Three Decades Post	-.035	.00021	.046	.013
	[0.031]	[0.051]	[0.035]	[0.001]
(K/L)Shock* Two Decades Post	-.034	-.0018	.024	-.028
	[0.023]	[0.021]	[0.035]	[0.007]
(K/L)Shock* One Decade Post	-.0065	.0016	.0032	-.0084
	[0.003]	[0.373]	[0.018]	[0.227]
N. obs	96	96	96	96
Y Mean	0.878	0.019	0.076	0.796
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.086	0.112	0.038	0.094
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.032	0.986	0.022	0.402
Panel B: Share of Men in Each Industry				
(K/L)Shock* Three Decades Post	.029	-.022	.0083	.071
	[0.015]	[0.001]	[0.016]	[0.196]
(K/L)Shock* Two Decades Post	-.0087	-.011	.023	.0058
	[0.024]	[0.017]	[0.008]	[0.025]
(K/L)Shock* One Decade Post	-.0043	-.0066	.0044	-.0026
	[0.000]	[0.030]	[0.005]	[0.237]
N. obs	96	96	96	96
Y Mean	0.728	0.087	0.152	0.577
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.088	0.222	0.232	0.018
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.572	0.118	0.154	0.080
Panel C: Testing female vs male				
$\alpha_{3,F} = \alpha_{3,M}$	0.004	0.000	0.644	0.002
$\alpha_{2,F} = \alpha_{2,M}$	0.936	0.004	0.982	0.728
$\alpha_{1,F} = \alpha_{1,M}$	0.044	0.006	0.018	0.774

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from May 1974 to Nov 1975/N. miners returning over the same time. Mean *K/L shock* = 0.73, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, trend interactions with a malaria dummy and two region dummies, and **trend interacted with indicators of above-or-below median value of each continuous control variable**.

Table A.6: Robustness: Weighting the K Shock by baseline district population

	Agriculture	Manufacturing	Services	Ind. Index
Panel A: Share of Women in Each Industry				
K Shock/Pop 66* Three Decades Post	-.19	.024	.22	.025
	[0.049]	[0.135]	[0.000]	[0.079]
K Shock/Pop 66* Two Decades Post	-.18	.01	.11	-.17
	[0.032]	[0.080]	[0.000]	[0.123]
K Shock/Pop 66* One Decade Post	-.028	.01	.0075	-.048
	[0.020]	[0.434]	[0.000]	[0.438]
N. obs	96	96	96	96
Y Mean	0.878	0.019	0.076	0.796
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.052	0.578	0.026	0.036
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.052	0.550	0.152	0.504
Panel B: Share of Men in Each Industry				
K Shock/Pop 66* Three Decades Post	.077	-.025	-.006	.19
	[0.098]	[0.002]	[0.000]	[0.429]
K Shock/Pop 66* Two Decades Post	-.065	.0015	.076	-.038
	[0.031]	[0.019]	[0.000]	[0.157]
K Shock/Pop 66* One Decade Post	-.033	-.011	.006	-.052
	[0.008]	[0.015]	[0.000]	[0.458]
N. obs	96	96	96	96
Y Mean	0.728	0.087	0.152	0.577
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.122	0.596	0.050	0.080
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.908	0.806	0.592	0.772
Panel C: Testing female vs male				
$\alpha_{3,F} = \alpha_{3,M}$ r.i pvalue	0.952	0.890	0.818	0.000
$\alpha_{2,F} = \alpha_{2,M}$ r.i pvalue	0.952	0.910	0.950	0.950
$\alpha_{1,F} = \alpha_{1,M}$ r.i pvalue	0.956	0.790	0.870	0.956

Randomization inference (r.i.) p-values in square brackets. ***K shock/Pop 66*** **is the deferred pay returning to each district between May 1974 and November 1975 in millions of USD divided by the district-level population in 1966.** Mean value is = 0.078, units are in 100USD. Unit of observation: district-gender cell. Controls: year and district fixed effects, interactions of a trend term with baseline district-level variables.

Table A.7: Robustness: Long-run impacts of K/L shock on sectoral employment shares, region-decade FE

	Agriculture	Manufacturing	Services	Ind. Index
Panel A: Share of Women in Each Industry				
(K/L)Shock* Three Decades Post	-.036	.005	.04	.00046
	[0.038]	[0.073]	[0.007]	[0.051]
(K/L)Shock* Two Decades Post	-.028	.0019	.017	-.028
	[0.023]	[0.033]	[0.007]	[0.135]
(K/L)Shock* One Decade Post	-.0036	.002	.00012	-.0083
	[0.013]	[0.379]	[0.000]	[0.498]
N. obs	96	96	96	96
Y Mean	0.878	0.019	0.076	0.796
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.066	0.424	0.018	0.320
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.090	0.406	0.136	0.408
Panel B: Share of Men in Each Industry				
(K/L)Shock* Three Decades Post	-.0026	.0028	.0091	.018
	[0.083]	[0.003]	[0.000]	[0.246]
(K/L)Shock* Two Decades Post	-.022	.0053	.021	-.018
	[0.029]	[0.015]	[0.000]	[0.045]
(K/L)Shock* One Decade Post	-.013	.0014	.0056	-.019
	[0.008]	[0.033]	[0.000]	[0.161]
N. obs	96	96	96	96
Y Mean	0.728	0.087	0.152	0.577
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.042	0.840	0.142	0.018
Test of sum: $\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.354	0.668	0.216	0.696
Panel C: Testing female vs male				
$\alpha_{3,F} = \alpha_{3,M}$ r.i pvalue	0.986	0.996	0.846	0.114
$\alpha_{2,F} = \alpha_{2,M}$ r.i pvalue	0.982	0.016	0.980	0.980
$\alpha_{1,F} = \alpha_{1,M}$ r.i pvalue	0.012	0.982	0.018	0.008

Randomization inference (r.i.) p-values in square brackets. *K/L shock* is district-level deferred pay from May 1974 to Nov 1975/N. miners returning over same time. Mean *K/L shock* = 0.73, units are in 100USD. Unit of observation: district-year cell. Controls: year and district fixed effects, linear trend interacted with all baseline district controls in Table 1, **two region dummies each interacted with decade dummies.**

Table A.8: Robustness of Long-run Impacts of K/L shock on Sectoral Employment Shares

	Agriculture [Mean = 0.88]				Services [Mean = 0.08]			
	Main Specification	Excluding trend controls	Excluding Lilongwe	Controlling for ΔMig_{67-73}	Main Specification	Excluding trend controls	Excluding Lilongwe	Controlling for ΔMig_{67-73}
Panel A: Share of Women								
(K/L)Shock* Three Decades Post	-.044	-.036	-.047	-.028	.049	.044	.053	.03
	[0.025]	[0.021]	[0.007]	[0.028]	[0.013]	[0.027]	[0.047]	[0.053]
(K/L)Shock* Two Decades Post	-.037	-.032	-.039	-.027	.025	.022	.027	.013
	[0.015]	[0.021]	[0.001]	[0.010]	[0.013]	[0.027]	[0.047]	[0.054]
(K/L)Shock* One Decade Post	-.009	-.0064	-.0099	-.004	.0052	.0038	.0064	-.00097
	[0.005]	[0.049]	[0.000]	[0.001]	[0.004]	[0.006]	[0.034]	[0.042]
N	96	96	92	96	96	96	92	96
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.104	0.058	0.022	0.090	0.104	0.070	0.028	0.064
$\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.020	0.038	0.002	0.026	0.036	0.028	0.004	0.016

	Agriculture [Mean = 0.73]				Services [Mean = 0.15]			
	Main Specification	Excluding trend controls	Excluding Lilongwe	Controlling for ΔMig_{67-73}	Main Specification	Excluding trend controls	Excluding Lilongwe	Controlling for ΔMig_{67-73}
Panel B: Share of Men								
(K/L)Shock* Three Decades Post	.0022	.016	.0025	.0063	.0063	.0082	.0091	-.0041
	[0.047]	[0.067]	[0.026]	[0.039]	[0.003]	[0.002]	[0.075]	[0.081]
(K/L)Shock* Two Decades Post	-.019	-.01	-.022	-.017	.018	.02	.021	.011
	[0.018]	[0.090]	[0.005]	[0.013]	[0.002]	[0.001]	[0.011]	[0.035]
(K/L)Shock* One Decade Post	-.011	-.0066	-.012	-.0097	.0038	.0044	.0055	.00023
	[0.004]	[0.115]	[0.003]	[0.002]	[0.001]	[0.000]	[0.058]	[0.050]
N	96	96	92	96	96	96	92	96
Tests:								
Joint test: $\alpha_1 = \alpha_2 = \alpha_3 = 0$	0.054	0.084	0.028	0.028	0.136	0.108	0.080	0.118
$\alpha_1 + \alpha_2 + \alpha_3 = 0$	0.446	0.960	0.398	0.582	0.236	0.078	0.126	0.596

Notes for Table A.8: Randomization inference p-values for each coefficient are in square brackets. Tests use RI p-values. K/L shock is deferred pay returning to each district between May 1974 and November 1975 in millions of USD divided by the number of miners returning to Malawi in those years. The variable is measured in units of 100USD per miner.Mean of K/L shock is 0.73. Outcomes data are from Census 1977 (baseline), 1987 (one decade post), 1998 (two decades post) and 2008 (three decades post). Unit of observation is the district-year cell. Total districts=24, except when Lilongwe is dropped. All regressions include year and district fixed effects. Except for columns (2) and (6), all regressions also include interactions of a time trend with baseline district controls (adult literacy in 1945, population density in 1945, a malaria dummy, share of men and women married in 1977, share of men and women not earning any cash income in 1966, tobacco and maize suitability indices, and two region dummies). Regressions in columns (4) and (8) additional control for interactions of year dummies with the number of migrants from each district returning between 1967 and 1973 (pre-plane crash).

ONLINE DATA APPENDIX: NOT FOR PUBLICATION

This online data appendix contains supplementary material for Dinkelman, Kumchulesi, and Mariotti. *"Labor migration, capital accumulation, and the structure of rural labor markets"*. A replication package (data and code) is available from the Harvard Dataverse.

A large group of overlapping generations of research assistants were instrumental in assembling some of the data for this project; most notably, the remittance data. We thank Freed Kumchulesi, Spencer Lambdin, Anwita Mahajan, Annelise Sauter-Ortiz, Mahnum Shazad, Dustin Sheehan, Khwima Singini, Ashley Wong, Lucy Xie, and Zheng-Yi Yang for their hard work in sourcing, cleaning, and organizing the data. We thank Regina Mannino and Anh Hoang for their excellent research assistance at the end stages of this project.

A Census data

Our main datasets are constructed from Census data collected in 1977, 1987, 1998 and 2008. The 1977 Census data were digitized from aggregate Census reports. The 100% microdata from the 1998 Census was obtained from the Malawi National Statistics Office. IPUMSI (`https://international.ipums.org/international/`) provides provides access to the 10% sample for 1998. The 1987 and 2008 Census data are 10% samples from the IPUMSI repository.

We also use data for some outcomes from earlier Census data in 1966, 1945 and 1931. We digitized all relevant tables from aggregate Census reports in these years (Malawi National Statistical Office, 1969; Nyasaland Governor, 1946, 1931). Datasets are available at `https://github.com/taryndinkelman/malawi/tree/main`

A.1 District boundary crosswalk: 1931 to 2008

We created a district boundary crosswalk that links district boundaries over time, through name changes and boundary changes. We use the districts existing in 1977 as the sample of districts. We consolidated information in variables from districts that had split in later years into their origin districts in 1977. For districts in earlier years that had split by the late 1970s, we apportioned the earlier cell totals to 1977 district boundaries using land area weights.

A.2 Labor market outcomes

We create three categories of labor market variables: broad sector of work variables, narrow sector of work variables, and economic activity status variables.

Broad sector of work: We define work in the agriculture, manufacturing, or service sector for each Census, using the number of people who are currently economically active (those employed and currently unemployed) in the denominator. Houseworkers and other inactive people (students, pensioners, other dependents) are excluded from both numerator and denominator of these variables. In each year, a small share of those in the labor force do not report an industry (most of these are unemployed people who have not worked before), so shares across the three broad sectors do not sum to one. For a more detailed definition of sector of work within the nonfarm sector, we disaggregate all non-agricultural employment into mining, manufacturing, retail, transport and communications, and all other services (business services, household services, and other non-specified services).

To create a summary measure of employment diversity in the district, we construct a Herfindahl index for (broad) industrial sector of work. The smaller the value of this index, the more evenly people are distributed across sectors. The larger the value of this index, the more people are concentrated within one of the three sectors.

Economic activity variables We define these variables for the sample that includes everyone in the relevant age group in a given district:

- In the labor force: working, unemployed, or doing home production
- Working: working or doing home production

- Subsistence: working as mlimi (subsistence farmer) or doing home production
- Family business worker: working in a non-farm family business
- Self-employed: working in a non-farm business for themselves
- Wage worker: working for someone else for a wage or salary
- Employer: employs other workers in a business

Economic activity variables and sector of work variables differ because the economic activity variables capture activity shares in the entire population, not just those in the labor force. Home production workers (mostly women) are excluded from sector of work variables but included in the economic activity variables. Our data show that the majority of family business workers, self-employed, wage workers and employers work in the non-farm sector.

In Data Appendix Tables A.1 and A.2, we compare the wording of Census questions across years. For the most part, it is possible to create a consistent set of definitions of each of the above variables, using combinations of different Census questions.

Data Appendix Table A.1: Occupation and Industry Questions in Malawi National Census

Census 1977	Census 1987	Census 1998	Census 2008
Sample: 10 years + answering yes to Qn. O	Sample: 10 years +, not inactive	Sample: 10 years and male, or female and not inactive (If inactive person is female, do not ask B18 and B19)	Sample: 10 years +, and ever worked (currently, or before) and currently available to work
Q: What is your occupation?	N: What is your occupation?	B18: What is this person's main occupation?	P25. What was [the respondent's] main occupation during the last 7 days or the last time he/she worked? P26. What is [the respondent's] status in the occupation? (Employer, self employed, public sector, private sector, family farm/business, other)
R: What is your industry of work?	O: What is your industry of work?	B19: What is this person's main trade or business (industry)?	P27. What is the main product, service or activity of [the respondent's] place of work?

A.3 Population density and urbanization variables

We digitized population data from the 1945, and 1966 Nyalasand Census and the 1977 Malawi Census. These data were reported at district level, sometimes separately for men and women in different age groups. We combined these data with district data from the 1987, 1998 and 2008 Census, and constructed population densities at district level using the area of the district. We also measure population totals, for men and women separately, and the share of population in urban areas within the district.

A.4 Migrants at district-level

In Census 1977, the total number of men who report ever migrating from Malawi is reported at district level (Census 1977, Table 4.8) while the share of miners who returned between 1966 and 1977 is reported in national aggregate data (Census 1977, Table 4.11). To construct district-specific numbers of migrants returning between 1966 and 1977, we multiplied the share of workers who had returned to Malawi in the last 10 years (out of all ever migrants who returned to Malawi) by the total men in each district who had ever migrated for work by 1977. Because of the labor ban, all migrants had returned to Malawi by 1975 and so would have been present in the 1977 Census.

A.5 Baseline district covariates from Census data

We used Census data to generate the following variables:

Historical literacy rates: we digitized data on the district-specific share of adults who were literate from the Report on the Census of 1931 (Nyasaland Protectorate, Table 6)

Share of married men and women in 1977: we digitized data on the share of men and women married from Census 1977 (Table 2.1)

Share of men and women with no cash incomes in 1966: we digitized the district-specific rates of men and women earning no cash income from the Malawi 1966 Population Census Final Report (Malawi National Statistics Office, Zomba: Table 21)

A.6 Physical and human capital investments and asset ownership

We measured investments in different ways, based on what information was available in at least two datasets. We used data from the 1977, 1987, 1998 and 2008 Census data as described above, and from the 1968/9 National Sample Survey of Agriculture (NSSA). The NSSA data were collected from around 5,000 households, and was designed to be representative at district-level. The part of the 1968 survey that collected these data was an income and expenditure-type survey.

Radios: The share of households in the district owning at least one radio exists in all years.

Durable housing: The share of households that lived in houses with a durable wall, durable roof, or both durable wall and roof was available in 1968 and in 1987.

Agricultural tools: The share of households with at least one panga, at least one hoe, or at least one type of livestock.

Education: The average level of education of individuals in specific age categories.

B DHS data

We use data from the 1992 Malawi Demographic Health Survey (DHS) to capture a larger set of household assets and to measure these assets 15 years after the end of the migration shock. The DHS data are publicly available upon registration from `https://dhsprogram.com/methodology/survey/survey-display-52.cfm`.

One of the key variables we use is the DHS wealth index, an index of assets constructed used principal components analysis. For more detail on how the variable is constructed, see `https://dhsprogram.com/programming/wealth%20index/DHS_Wealth_Index_Files.pdf`

C Administrative data

To measure flows of migrant capital, and describe the composition of miners, we collected and digitized data from the National Archives in Malawi and from The Employment Bureau of Africa (TEBA) archives in South Africa, from the Malawian National archives and Rhodes House Library at Oxford University in the U.K.

Migrant capital: Our data record the monthly flows of migrant money from South Africa to specific districts in Malawi, for the period October 1966 to November 1975. These records come from documents entitled "Attestation and Despatch Returns to the Ministry of Labour", found in Malawi's National Archives in Zomba and in the TEBA Archives at the University of Johannesburg, South Africa. To construct a time series of the flows in a consistent currency unit, we converted GBP to the Malawi Kwacha using an exchange rate of 2:1, the official exchange rate at the time the Malawi currency was adopted in 1971. Capital flows were recorded in each of three categories: deferred pay, voluntary remittances, and deposits. Our analysis uses only the deferred pay amounts that were set by contract. These flows make up 89% of the total flows of money over the period.

D Other Geographic Data

We also created the following variables:

Area: geographic area for 24 districts was calculated in ArcGIS.

High Malaria Area indicator: we computed altitude for each point on the Malawian grid map using data from the national map seamless server `http://seamless.usgs.gov/index.php` and the Viewshed tool in ArcGIS. We aggregated these measures to district level. Then we defined areas of high, medium or low malaria susceptibility based on standard measures of altitude: high malaria areas (altitude below 650m), medium malaria areas (altitudes between 650m and 1100m) and low malaria areas (altitudes over 1100m)

Estate indicator: We identified which districts contained a large tea or tobacco plantation using information in Christiansen (1984). The FAO's crop suitability index measuring whether a district is highly suitable for tobacco or tea production significantly predicts this estate district indicator

Crop suitability indices: Maize and tobacco suitability indices were constructed using GAEZ FAO data (`https://gaez.fao.org/pages/data-access-download`) and QGIS software. We selected the Suitability Index Range data from 1961-1990 with the following criteria: low input level, rainfed, no CO2 Fertilization, all land in grid cell. This was transferred into QGIS, combined with a shape file containing Malawi district boundaries, and used to calculate the mean suitability index in each municipality.

References

Chirwa, Wiseman Chijere. 1992."TEBA is power: Rural labour, migrancy and fishing in Malawi, 1890s – 1985", Ph.D thesis, Queen's University, Kingston, Ontario.

Crush, Jonathan. S., Alan Jeeves and David Yudelman. 1991. South Africa's Labor Empire. Westview Press.

Malawi National Statistical Office. 1970. *National Sample Survey of Agriculture 1968 and 1969: Customary land in rural areas only.* Zomba

Malawi National Statistical Office. 1984. *National Sample Survey of Agriculture 1980 and 1981: Customary land in rural areas only.*, Zomba

Malawi National Statistical Office. 1996. *National Sample Survey of Agriculture 1992 and 1993: Vol. 1, 2, 3, 4.*, Zomba

Malawi National Statistical Office. 1991. *Malawi 1987 Population and Housing Census Summary of Final Reports.*, Volumes I and II, Zomba

Malawi National Statistical Office. 1979. *Malawi 1977 Population Census Final Report*, Volumes I and II, Zomba

Malawi National Statistical Office. 1967. *Malawi 1966 Population Census Final Report*, Government Printer, Zomba

Nyasaland Governor. 1956. "Memorandum on Labor Migrancy in Malawi." 1956 Nyasaland Provincial Office Memo, December 7 1961.

Nyasaland Protectorate. 1931. *Report on the Census of 1931.* Government Printer, Zomba

Nyasaland Protectorate. 1946. *Report on the Census of 1945.* Government Printer, Zomba

Wilson, Francis. 1972. Labour in the South African Gold Mines 1911-1969. Cambridge University Press.