ONLINE RESULTS APPENDIX: NOT FOR PUBLICATION

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

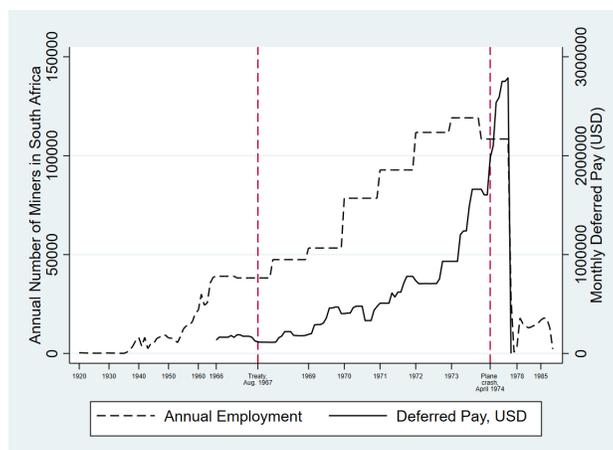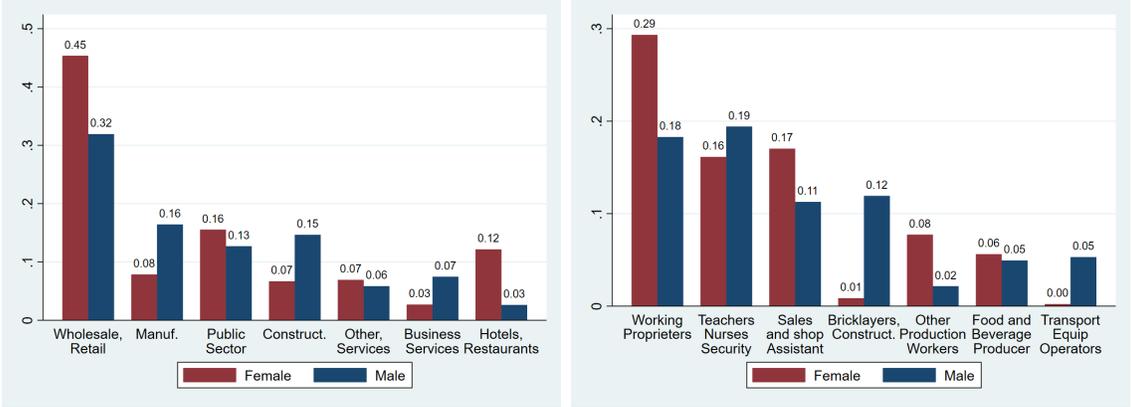# Figure A.1: Capital Flows and Migrant Workers Over Time

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.

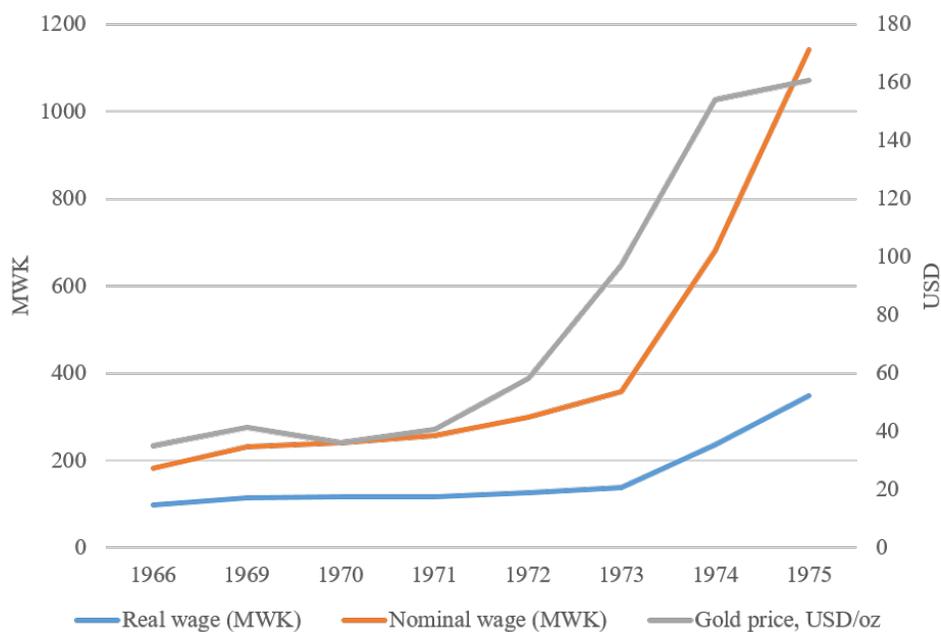**Figure A.3: Annual Wages for African Mineworkers and Global Gold Prices**

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.**

8

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**.

9

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 supplmentary 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 for her 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 planta-

tion 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.