In [1]:

```
root_path = 'C:/Users/rodri/Dropbox/Malawi/SIEG2021 (1)/2023 July'
folder_fig = root_path+'/Figures'

import numpy as np
import pandas as pd
import os

import warnings
# Suppress warnings in the entire notebook
warnings.filterwarnings("ignore")

pd.options.display.max_columns = None
pd.options.display.max_rows = None

os.chdir('C:/Users/rodri/Dropbox/JMP/python')
from data_functions_albert import remove_outliers, gini
os.chdir(root_path)

save=True



dollar_MWK = 1030.36
pd.options.display.float_format = '{:,.2f}'.format
percentiles=[0.5]

#income
inc = pd.read_csv(root_path+'/Data/Clean data/Phase 3 - Consumption, Transfers, Inco
#consumption
cons = pd.read_csv(root_path+'/Data/Clean data/Phase 3 - Consumption, Transfers, Inc
#Roster
roster = pd.read_csv(root_path+'/Data/Clean data/Phase 1 - Roster/roster_jul23.csv')

foodtra = pd.read_csv(root_path+'/Data/Clean data/Phase 3 - Consumption, Transfers,



data = pd.merge( inc, roster, on='hhid', how='left')
data = pd.merge(data, cons, on='hhid', how='inner')
data = pd.merge(data, foodtra, on='hhid', how='left')

del data['cashtrans_yes.1'], data['k_farm.1']


data.rename(columns={'hh_area_plots': 'land_area', 'hh_value_plots': 'land_value'},




data['inctotal_cap'] = data['inctotal']/data['hh_size'].replace(0,np.mean(data['hh_s
data['ctotal_cap'] = data['ctotal']/data['hh_size'].replace(0,np.mean(data['hh_size'
data['inctotal_trans_cap'] = data['inctotal_trans']/data['hh_size'].replace(0,np.mea
data['land_area_cap'] = data['land_area']/data['hh_size'].replace(0,np.mean(data['hh
data['y_net_cap'] = data['y_net']/data['hh_size'].replace(0,np.mean(data['hh_size'])

data['ln_inc'] =  np.log(data['inctotal']).replace([-np.inf, np.inf], np.nan)
data['ln_c'] = np.log(data['ctotal']).replace([-np.inf, np.inf], np.nan)
data['ln_land'] = np.log(data['land_area']).replace([-np.inf, np.inf], np.nan)
data['ln_inctrans'] =  np.log(data['inctotal_trans']).replace([-np.inf, np.inf], np.
data['ln_agric'] =  np.log(data['y_net']).replace([-np.inf, np.inf], np.nan)
```

```
data['ln_inc_cap'] =  np.log(data['inctotal_cap']).replace([-np.inf, np.inf], np.nan
data['ln_c_cap'] = np.log(data['ctotal_cap']).replace([-np.inf, np.inf], np.nan)
data['ln_land_cap'] = np.log(data['land_area_cap']).replace([-np.inf, np.inf], np.na
data['ln_inctrans_cap'] =  np.log(data['inctotal_trans_cap']).replace([-np.inf, np.i
data['ln_agric_cap'] =  np.log(data['y_net_cap']).replace([-np.inf, np.inf], np.nan)


data['rank_inctotal'] = data['inctotal'].rank(pct=True)
data['rank_landarea'] = data['land_area'].rank(pct=True)
data['rank_landvalue'] = data['land_value'].rank(pct=True)
data['rank_ctotal'] = data['ctotal'].rank(pct=True)


data['rank_wtotal'] = data['wtotal'].rank(pct=True)
data['wtotal_cap'] = data['wtotal']/data['hh_size']
data['ln_w'] = np.log(data['wtotal']+np.abs(np.min(data['wtotal']))).replace([-np.in
data['ln_w_cap'] = np.log(data['wtotal_cap']+np.abs(np.min(data['wtotal']))).replace



print( '                ')
print('================================================================')
print( 'HOUSEHOLD DATASET WAVE JULY 2023')
print('merges roster, income_weath, consumption, and hhtransfers datasets')
print( '                ')
print('================================================================')
if save==True:
    data.to_csv(root_path+'/Data/Clean data/hhdata23.csv', index=False)

print('Data saved: hhdata23.csv')
print('Dataset contains the following variables:')
print(data.columns.tolist())



### SUMMARY

print( '                ')
print('================================================================')
print(' SOCIODEMOGRAPHIC CHARACTERISTICS')
print('================================================================')
print( '                ')
print(data[['hh_size', 'head_gender', 'head_marital', 'head_age','interviewed_feb23'



## Summary CIW 2019
data[['inctotal']] = data[['inctotal']].replace([0, 0.00], np.nan)

data_sum = data[['ctotal','inctotal','wtotal','land_area','ctotal_cap','inctotal_cap

data_sum[['ctotal','inctotal','wtotal','ctotal_cap','inctotal_cap','wtotal_cap']] =

sum_cwi = pd.DataFrame((data_sum[['ctotal','inctotal','wtotal','land_area']]).mean(a
sum_cwi = sum_cwi.append(pd.DataFrame((data_sum[['ctotal','inctotal','wtotal','land_

varlog_df = pd.DataFrame(((data_sum[['ln_c','ln_inc','ln_w','ln_land']]).var(axis=0)

sum_cwi = sum_cwi.append(dict(zip(sum_cwi.columns, np.array(varlog_df.iloc[:,0]).T))


print( '                ')
print('================================================================')
```

```
print(' CONSUMPTION. INCOME, AND WEALTH')
print('================================================================')
print( '                 ')
print(sum_cwi)

sum_cwi_cap = pd.DataFrame((data_sum[['ctotal_cap','inctotal_cap','wtotal_cap','land
sum_cwi_cap = sum_cwi_cap.append(pd.DataFrame((data_sum[['ctotal_cap','inctotal_cap'

varlog_df = pd.DataFrame(((data_sum[['ln_c_cap','ln_inc_cap','ln_w_cap','ln_land_cap

sum_cwi_cap = sum_cwi_cap.append(dict(zip(sum_cwi_cap.columns, np.array(varlog_df.il

print(sum_cwi_cap)
print('Mean, median, and log-variance')

# In per capita terms



### Summary I
income = data[['inctotal','y_net', 'wlabor_inc', 'ganyu_inc', 'business_profits', 'o


sum_inc = (income.loc[:, income.columns !='hhid']/dollar_MWK).describe(percentiles=[

obs_inc = sum_inc.iloc[0,:]
shares_employ = obs_inc/238

shares_gdp = data[['inctotal','y_net', 'wlabor_inc', 'ganyu_inc', 'business_profits'

shares_gdp = shares_gdp/shares_gdp[0]

shares =  pd.concat([shares_gdp, shares_employ], axis=1).T


print( '                 ')
print('================================================================')
print(' INCOME AND EMPLOYMENT SHARES')
print('================================================================')
print( '                 ')
print(shares)


## no one grew: cassava, sugarcane,pearlmillet

sum_agric =  ((data[['y_agric','y_maize', 'y_groundnut', 'y_groundbean', 'y_sweetpot

obs_agric = sum_agric.iloc[0,:]
shares_crops = obs_agric/238


shares_agric = data[['y_agric','y_maize', 'y_groundnut', 'y_groundbean', 'y_sweetpot

shares_agric = shares_agric/shares_agric[0]

shares =  pd.concat([shares_agric, shares_crops], axis=1).T


print( '                 ')
print('================================================================')
print(' AGRICULTURAL SHARES')
print('================================================================')
print( '                 ')
print(shares)
```

```
###Summary C

print( '                 ')
print('=================================================================')
print(' CONSUMPTION')
print('=================================================================')
print( '                 ')


c_summary =  ((data[['ctotal','c_food','c_food_purch','c_food_ownprod', 'c_nonfood',
print(c_summary)



print( '                 ')
print('=================================================================')
print(' WEALTH')
print('=================================================================')
print( '                 ')

### Summary W
wealth = data[['wtotal','land_value','k_farm', 'hhlivestock', 'housing', 'hh_assets'


sum_w = wealth.describe(percentiles=[0.5])

var_list = ['wtotal','land_value','k_farm', 'hhlivestock', 'housing', 'hh_assets']
gini_stat= np.empty((1, len(var_list)))

for i,state in enumerate(var_list):
    gini_stat[:,i] = gini(wealth[state].dropna().values)
data_gini = pd.DataFrame(gini_stat, columns=var_list)
data_gini.reset_index(inplace=True)
data_gini['index'] = 'gini'
sum_w.reset_index(inplace=True)
sum_w = sum_w.append(data_gini, ignore_index=True)
print(sum_w)
```

```
============================================================
HOUSEHOLD DATASET WAVE JULY 2023
merges roster, income_weath, consumption, and hhtransfers datasets

============================================================
Data saved: hhdata23.csv
Dataset contains the following variables:
['hhid', 'rightsellland', 'chiefpreventsell', 'chiefpreventbequeat', 'cashtrans_ye
s', 'govcoupon', 'inctotal', 'inctotal_trans', 'y_net', 'y_agric', 'y_maize', 'y_gro
undnut', 'y_pigeonpeas', 'total_kg_maize', 'total_kg_groundnut', 'total_kg_pigeonpea
s', 'y_cassava', 'y_soyabean', 'y_sorghum', 'y_fingermillet', 'y_cotton', 'y_tanapos
i', 'y_groundbean', 'y_nkhwani', 'y_sugarcane', 'y_sweetpotatoe', 'sold_agric', 'sol
d_insiders_agric', 'store_agric', 'land_area', 'hh_ratio_value_rent', 'hh_p_acre_plo
ts', 'area_cultivated', 'k_farm', 'labor_N', 'labor_h', 'hh_labor_hours', 'hired_men
_L', 'hired_women_L', 'hired_kids_L', 'interm', 'fertilizerkg', 'p_fert', 'value_fer
tilizer', 'spendseeds', 'spendpesticides', 'wlabor_inc', 'wlabor_supply', 'ganyu_ye
s', 'ganyu_inc', 'ganyu_supply', 'business_revenue', 'business_costs', 'business_pro
fits', 'business_profits2', 'NGO_yes', 'gov_yes', 'remittances_yes', 'other_inc', 'c
ashtrans_value', 'NGO_trans', 'gov_trans', 'remittances', 'wtotal', 'housing', 'hh_a
ssets', 'land_value', 'hhlivestock', 'shocks', 'shock_flood', 'shock_drought', 'shoc
k_lndslide', 'shock_covid', 'shock_adultill', 'shock_kidill', 'shock_death_earner',
'shock_death_othermemb', 'shock_inp_p', 'shock_out_p', 'shock_pests', 'shock_lvstk',
```

```
'shock_theft', 'shock_theft_agric', 'shock_business', 'shock_unemp', 'shock_wage_dec
r', 'shock_other', 'wave', 'invillage_feb23', 'interviewed_feb23', 'intervieweenam
e', 'head_name', 'village', 'subvillage', 'key_landmark', 'mosque_church', 'hh_siz
e', 'hh_phone', 'head_gender', 'head_marital', 'head_age', 'head_nickname', 'head_ed
uc', 'head_religion', 'head_female', 'head_married_mono', 'head_married_poly', 'head
_nevermarried', 'head_divorced', 'head_widowed', 'head_separated', 'head_christian',
'head_noeduc', 'spouse_educ', 'ethnic', 'mlanguage', 'village_born', 'village_year
s', 'chief_related', 'chief_relation', 'elder_yes', 'elders_related', 'elders_relati
on', 'head_belowprimary4', 'head_belowprimary7', 'head_belowsecond3', 'head_secondar
y', 'head_educ_countin', 'gps_lat_3', 'gps_long_3', 'c_food', 'c_food_purch', 'c_foo
d_ownprod', 'c_nonfood', 'c_housing', 'c_clothes', 'c_education', 'c_health', 'c_fun
eralout', 'c_funeralin', 'c_weddingout', 'c_weddingin', 'ctotal', 'transfers1_net',
'transfers2_net', 'transfers3_net', 'inctotal_cap', 'ctotal_cap', 'inctotal_trans_ca
p', 'land_area_cap', 'y_net_cap', 'ln_inc', 'ln_c', 'ln_land', 'ln_inctrans', 'ln_ag
ric', 'ln_inc_cap', 'ln_c_cap', 'ln_land_cap', 'ln_inctrans_cap', 'ln_agric_cap', 'r
ank_inctotal', 'rank_landarea', 'rank_landvalue', 'rank_ctotal', 'rank_wtotal', 'wto
tal_cap', 'ln_w', 'ln_w_cap']

============================================================
 SOCIODEMOGRAPHIC CHARACTERISTICS
============================================================

       hh_size  head_gender  head_marital  head_age  head_married_mono  \
count   283.00       283.00        283.00    283.00             284.00
mean      4.84         1.31          2.02     43.64               0.63
std       1.99         0.46          1.56     17.69               0.48
min       1.00         1.00          1.00     11.00               0.00
25%       3.00         1.00          1.00     28.00               0.00
50%       4.00         1.00          1.00     41.00               1.00
75%       6.00         2.00          3.00     53.50               1.00
max      11.00         2.00          6.00     91.00               1.00

       head_married_poly  head_divorced  head_belowprimary4  \
count             284.00         284.00              284.00
mean                0.09           0.17                0.18
std                 0.29           0.38                0.39
min                 0.00           0.00                0.00
25%                 0.00           0.00                0.00
50%                 0.00           0.00                0.00
75%                 0.00           0.00                0.00
max                 1.00           1.00                1.00

       head_belowprimary7
count              284.00
mean                0.28
std                 0.45
min                 0.00
25%                 0.00
50%                 0.00
75%                 1.00
max                 1.00

============================================================
 CONSUMPTION. INCOME, AND WEALTH
============================================================

    ctotal  inctotal   wtotal  land_area
0 1,318.33    403.26 1,410.40       2.11
1 1,117.47    261.02   873.48       2.00
2     0.34      1.16     1.31       0.47
   ctotal_cap  inctotal_cap  wtotal_cap  land_area_cap
0      303.99         89.47      310.39           0.46
1      236.41         48.54      181.09           0.36
2        0.35          1.17        1.18           0.40
Mean, median, and log-variance

============================================================
 INCOME AND EMPLOYMENT SHARES
============================================================
```

	inctotal	y_net	wlabor_inc	ganyu_inc	business_profits	other_inc
0	1.00	0.62	0.05	0.13	0.19	0.19
count	1.15	1.12	0.03	0.29	0.26	0.54

==========
AGRICULTURAL SHARES
==========

	y_agric	y_maize	y_groundnut	y_groundbean	y_sweetpotatoe	\
0	1.00	0.52	0.26	0.01	0.03	
count	1.12	1.05	0.55	0.16	0.13	

	y_pigeonpeas	y_nkhwani
0	0.11	0.02
count	1.01	0.21

==========
CONSUMPTION
==========

	ctotal	c_food	c_food_purch	c_food_ownprod	c_nonfood	c_housing	\
count	284.00	284.00	283.00	276.00	284.00	282.00	
mean	1,318.33	1,034.27	556.80	440.78	284.06	185.99	
std	848.65	620.65	435.37	524.31	426.91	241.43	
min	263.94	215.41	4.66	1.55	5.82	5.82	
50%	1,117.47	873.27	431.85	294.53	176.25	116.46	
max	6,838.62	3,710.38	2,789.14	3,730.60	4,760.67	2,911.60	

	c_clothes	c_education	c_health
count	194.00	163.00	203.00
mean	74.85	50.22	24.77
std	135.95	218.13	43.25
min	1.55	0.78	0.78
50%	34.94	15.53	11.65
max	1,552.86	2,601.03	388.21

==========
WEALTH
==========

	index	wtotal	land_value	k_farm	hhlivestock	housing	hh_assets
0	count	284.00	284.00	284.00	284.00	284.00	284.00
1	mean	1,410.40	663.56	18.85	51.35	588.40	88.23
2	std	1,542.74	958.76	44.37	137.84	847.67	178.04
3	min	0.00	0.00	0.00	0.00	0.00	0.00
4	50%	873.48	388.21	9.22	9.71	291.16	25.72
5	max	11,500.16	5,823.21	609.50	1,989.60	7,764.28	1,487.83
6	gini	0.51	0.61	0.62	0.77	0.61	0.73

In [ ]: