In [2]:

```
root_path = 'C:/Users/rodri/Dropbox/Malawi/SIEG2021 (1)/2022 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=False



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_22.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 2022')
print('merges roster, income_weath, consumption, and hhtransfers datasets')
print( '                ')
print('================================================================')
if save==True:
    data.to_csv(root_path+'/Data/Clean data/hhdata22.csv', index=False)

print('Data saved: hhdata22.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_19','h



## 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 2022
merges roster, income_weath, consumption, and hhtransfers datasets

================================================================
Data saved: hhdata22.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_19', 'interviewed_19', 'oldhhid', 'interviewee
name', 'head_name', 'village', 'subvillage', 'key_landmark', 'mosque_church', 'hh_si
ze', 'hh_phone', 'head_gender', 'head_marital', 'head_age', 'head_nickname', 'head_e
duc', 'head_religion', 'head_female', 'head_married_mono', 'head_married_poly', 'hea
d_nevermarried', 'head_divorced', 'head_widowed', 'head_separated', 'head_christia
n', 'head_noeduc', 'spouse_educ', 'ethnic', 'mlanguage', 'village_born', 'village_ye
ars', 'chief_related', 'chief_relation', 'elder_yes', 'elders_related', 'elders_rela
tion', 'head_belowprimary4', 'head_belowprimary7', 'head_belowsecond3', 'head_second
ary', 'head_educ_countin', 'gps_lat', 'gps_long', 'c_food', 'c_food_purch', 'c_food_
ownprod', 'c_nonfood', 'c_housing', 'c_clothes', 'c_education', 'c_health', 'c_funer
alout', 'c_funeralin', 'c_weddingout', 'c_weddingin', 'ctotal', 'transfers1_net', 't
ransfers2_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   272.00       272.00        272.00    272.00             272.00
mean      4.47         1.35          2.06     43.51               0.62
std       1.87         0.48          1.56     17.58               0.49
min       1.00         1.00          1.00     10.00               0.00
25%       3.00         1.00          1.00     29.00               0.00
50%       4.00         1.00          1.00     40.50               1.00
75%       6.00         2.00          4.00     53.25               1.00
max      11.00         2.00          6.00     90.00               1.00

       head_married_poly  head_divorced  head_belowprimary4  \
count             272.00         272.00              272.00
mean                0.09           0.19                0.18
std                 0.28           0.40                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              272.00
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  872.61    297.89 1,160.30       2.21
1  764.34    148.93   863.29       1.50
2    0.43      1.48     0.55       0.53
   ctotal_cap  inctotal_cap  wtotal_cap  land_area_cap
0      214.02         76.45      301.22           0.57
1      173.88         34.80      186.83           0.38
2        0.40          1.51        0.38           0.57
Mean, median, and log-variance

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

       inctotal  y_net  wlabor_inc  ganyu_inc  business_profits  other_inc
0          1.00   0.30        0.05       0.42              0.22       0.22
count      1.12   1.06        0.03       0.43              0.26       0.55

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

       y_agric  y_maize  y_groundnut  y_groundbean  y_sweetpotatoe  \
0         1.00     0.74         0.15          0.01            0.00
count     1.06     1.06         0.40          0.05            0.02

       y_pigeonpeas  y_nkhwani
0              0.05       0.01
```

```
count          0.62       0.03

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

        ctotal   c_food  c_food_purch  c_food_ownprod  c_nonfood  c_housing  \
count   272.00   272.00        272.00          258.00     272.00     270.00
mean    872.61   656.71        389.03          208.61     215.91     143.97
std     541.11   386.61        268.96          254.75     271.36     172.86
min     109.45    98.38         21.01            1.73       0.78       2.91
50%     764.34   598.87        322.04          117.07     120.54      78.03
max   3,531.70 2,460.22      1,682.23        1,639.92   2,042.00     931.71

       c_clothes  c_education  c_health
count     141.00       149.00    202.00
mean       61.51        33.68     23.33
std       117.53        83.72     42.41
min         0.78         0.39      0.19
50%        19.41        13.59      7.96
max       870.76       854.07    388.21

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

   index    wtotal  land_value  k_farm  hhlivestock   housing  hh_assets
0  count    272.00      272.00  272.00       272.00    272.00     272.00
1   mean  1,160.30      446.36   13.49        40.31    607.88      52.26
2    std  1,641.60      622.76   26.61        73.35  1,339.76     100.31
3    min     67.94        0.00    0.00         0.00      0.00       0.00
4    50%    863.29      291.16    7.28         6.79    291.16      19.41
5    max 18,872.04    5,726.15  284.85       510.02 14,558.02     732.75
6   gini      0.46        0.52    0.60         0.75      0.59       0.72
```

In [ ]: