In [1]:
```python
# ==========================================================================
#  VILLAGE PANEL --- 2018, 2019, 2022, 2023
# ==========================================================================

import numpy as np
import pandas as pd
import os

os.chdir('C:/Users/rodri/Dropbox/JMP/python')
from data_functions_albert import remove_outliers, gini
pd.options.display.float_format = '{:,.2f}'.format

dollar_MWK = 1030.36

## Display  set-up
pd.options.display.float_format = '{:,.2f}'.format
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

root_path = 'C:/Users/rodri/Dropbox/Malawi/'
folder_fig = root_path+'/Figures'


data1819 = pd.read_csv('C:/Users/rodri/Dropbox/Malawi/Chied_Field_June_19/Data/Finis
data22 = pd.read_csv('C:/Users/rodri/Dropbox/Malawi/SIEG2021 (1)/2022 July/Data/Clea
data23 = pd.read_csv('C:/Users/rodri/Dropbox/Malawi/SIEG2021 (1)/2023 July/Data/Clea


# print variables 1819


list_vars1 = data1819.columns.to_list()
list_vars2 = data22.columns.to_list()
unique_to_list1 = [item for item in list_vars1 if item not in list_vars2]

# Find elements in list2 that are not in list1
unique_to_list2 = [item for item in list_vars2 if item not in list_vars1]

#print(unique_to_list1)
#print(unique_to_list2)

save_data=False



data23['wave']=2023

print('cleaning and preparing the panel')
#make variables to coincide across waves. Delete some variables only asked in one wa

data1819.rename(columns={'hhid':'oldhhid',  'hh_head_phone':'head_phone',
                         'spouse_educ_countin':'spouse_educ','villageheadfamily':'ch

data1819 = data1819.drop(columns=['k_farm_18', 'hhlivestock_18' ,'housing_18', 'hh_a
                         'wlabor_yes', 'business_yes', 'net_degree', 'net_between', '
                         'rightsellplot', 'denominationhead'])

data22 = data22.drop(columns=[ 'head_nevermarried','head_nickname','spendseeds', 'sp

data23 = data23.drop(columns=[ 'head_nevermarried','head_nickname','spendseeds', 'sp
```

```python
### fix old vs new hhid

data1819 = data1819.merge(data22[['hhid','oldhhid']],on='oldhhid', how='left')

check = data1819[['hhid','oldhhid','wave']]

data1819['hhid'].fillna(data1819['oldhhid'], inplace=True)


### recategorize head_marital
data1819['head_marital'].unique()

pd.value_counts(data22['head_marital'])

data22['head_marital'].replace([1.0,2.0,3.0,4.0,5.0,6.0],['monogamous married', 'pol
data23['head_marital'].replace([1.0,2.0,3.0,4.0,5.0,6.0],['monogamous married', 'pol


duplicate_columns = data1819.columns[data1819.columns.duplicated()]

## Construct panel.

print( '                     ')
print('================================================================')
print(' PANEL CHECK')
print('================================================================')

data = data22.append(data23)
data = data.append(data1819)
print('Number of households in each survey')
print(pd.value_counts(data['wave']))

print('Number of households observed in repeated waves:')
print(pd.value_counts(data.groupby(by="hhid")["hhid"].count()))
print('160 households were in the 4 waves of data')

print('Number of households in both 2022 and 2023 waves:',len(data22.merge(data23,on

print('Number of households in  2019, 2022, and 2023 waves:',len((data22.merge(data2


print( '                     ')
print('================================================================')
print(' SUMMARY STATISTICS ACROSS WAVES')
print('================================================================')
print('Averages: Consumption, Income, Wealth')

print(data[['ctotal','inctotal','wtotal', 'wave']].groupby(by='wave').mean()/dollar_
print('in dollars, exchange rate july 2023. Not controlled for inflation')

print( '                     ')
print('================================================================')
print('Inequality:variance of the log of consumption, income, wealth, and land size'
print(data[['ln_c','ln_inc','ln_w', 'ln_land', 'wave']].groupby(by='wave').var())

print( '                     ')
print('================================================================')
print('Household characeristics')
print(data[['hh_size','land_area','head_age','head_female','head_divorced','wave']].

print( '                     ')
print('================================================================')
print('Agriculture')
print(data[['total_kg_maize', 'total_kg_groundnut',  'total_kg_pigeonpeas','land_are
```

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

print( '                    ')
print('============================================================')
print('Non-village transfers: Ngo,gov, remittances')
print(data[['NGO_trans', 'gov_trans', 'remittances','wave']].groupby(by='wave').mean
print('============================================================')

if save_data==True:
    data.to_csv('C:/Users/rodri/Dropbox/Malawi\SIEG2021 (1)/2023 July/Data/Clean dat

print( '                  ')

print( '                  ')
print(' HOUSEHOLD PANEL SAVED: 2023 July/Clean data/panel_village_18to23.csv ')
print('Variables in the panel:')
print(data.columns.to_list())
```

cleaning and preparing the panel

```
============================================================
 PANEL CHECK
============================================================
Number of households in each survey
2023    284
2022    272
2018    269
2019    246
Name: wave, dtype: int64
Number of households observed in repeated waves:
4    160
2    102
1    102
3     40
5      1
Name: hhid, dtype: int64
160 households were in the 4 waves of data
Number of households in both 2022 and 2023 waves 244
Number of households in  2019, 2022, and 2023 waves 170


============================================================
 SUMMARY STATISTICS ACROSS WAVES
============================================================
Averages: Consumption, Income, Wealth
      ctotal   inctotal   wtotal
wave
2018   341.41    246.39 1,019.25
2019   303.30    273.30   924.04
2022   872.61    292.42 1,160.30
2023 1,318.33    389.06 1,410.40
in dollars, exchange rate july 2023. Not controlled for inflation


============================================================
Inequality:variance of the log of consumption, income, wealth, and land size
      ln_c   ln_inc   ln_w   ln_land
wave
2018  0.66    2.54   1.45      0.47
2019  0.29    0.95   1.11      0.54
2022  0.43    1.48   0.55      0.53
2023  0.34    1.16   1.31      0.47


============================================================
Household characeristics
      hh_size   land_area   head_age   head_female   head_divorced
wave
2018    4.56      2.33        53.10          nan             nan
2019    4.48      2.62        47.24         0.00            0.13
2022    4.47      2.21        43.51         0.35            0.19
```

| 2023 | 4.84 | 2.11 | 43.64 | 0.31 | 0.17 |

```
===============================================================
```
Agriculture

	total_kg_maize	total_kg_groundnut	total_kg_pigeonpeas	land_area \
wave				
2018	441.75	258.53	295.58	2.33
2019	294.23	99.98	18.59	2.62
2022	230.06	14.39	9.81	2.21
2023	187.22	64.41	49.90	2.11

	fertilizerkg	hh_labor_hours
wave		
2018	59.45	828.38
2019	49.75	1,411.66
2022	65.40	1,015.66
2023	36.72	770.11

```
===============================================================
```

```
===============================================================
```
Non-village transfers: Ngo,gov, remittances

	NGO_trans	gov_trans	remittances
wave			
2018	nan	2.27	5.92
2019	8.71	2.26	15.14
2022	6.00	0.95	55.81
2023	4.79	1.49	37.82

```
===============================================================
```

 HOUSEHOLD PANEL SAVED: 2023 July/Clean data/panel_village_18to23.csv
Variables in the panel:
['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', 'wlabor_inc', 'wlabor_supply', 'ganyu_yes', 'ganyu_inc', 'ganyu_supply',
'business_revenue', 'business_costs', 'business_profits', 'business_profits2', 'NGO_
yes', 'gov_yes', 'remittances_yes', 'other_inc', 'cashtrans_value', 'NGO_trans', 'go
v_trans', 'remittances', 'wtotal', 'housing', 'hh_assets', 'land_value', 'hhlivestoc
k', 'shocks', 'shock_flood', 'shock_drought', 'shock_lndslide', 'shock_covid', 'shoc
k_adultill', 'shock_kidill', 'shock_death_earner', 'shock_death_othermemb', 'shock_i
np_p', 'shock_out_p', 'shock_pests', 'shock_lvstk', 'shock_theft', 'shock_theft_agri
c', 'shock_business', 'shock_unemp', 'shock_wage_decr', 'shock_other', 'wave', 'invi
llage_19', 'interviewed_19', 'oldhhid', 'intervieweename', 'head_name', 'village',
'subvillage', 'key_landmark', 'mosque_church', 'hh_size', 'hh_phone', 'head_gender',
'head_marital', 'head_age', 'head_educ', 'head_religion', 'head_female', 'head_marri
ed_mono', 'head_married_poly', 'head_divorced', 'head_widowed', 'head_separated', 'h
ead_christian', 'head_noeduc', 'spouse_educ', 'ethnic', 'mlanguage', 'village_born',
'village_years', 'chief_related', 'chief_relation', 'elder_yes', 'elders_related',
'elders_relation', 'head_belowprimary4', 'head_belowprimary7', 'head_belowsecond3',
'head_secondary', 'head_educ_countin', 'gps_lat', 'gps_long', 'c_food', 'c_food_purc
h', 'c_food_ownprod', 'c_nonfood', 'c_housing', 'c_clothes', 'c_education', 'c_healt
h', 'c_funeralout', 'c_funeralin', 'c_weddingout', 'c_weddingin', 'ctotal', 'transfe
rs1_net', 'transfers2_net', 'transfers3_net', 'inctotal_cap', 'ctotal_cap', 'inctota
l_trans_cap', 'land_area_cap', 'y_net_cap', 'ln_inc', 'ln_c', 'ln_land', 'ln_inctran
s', 'ln_agric', 'ln_inc_cap', 'ln_c_cap', 'ln_land_cap', 'ln_inctrans_cap', 'ln_agri
c_cap', 'rank_inctotal', 'rank_landarea', 'rank_landvalue', 'rank_ctotal', 'rank_wto
tal', 'wtotal_cap', 'ln_w', 'ln_w_cap', 'invillage_feb23', 'interviewed_feb23', 'gps
_lat_3', 'gps_long_3', 'rightbequeathplot', 'hhlabor_N', 'hired_N', 'hired_labor_hou
rs', 'kg_fertilizer', 'shock_dead', 'y_therereokra', 'y_tomatoes', 'y_pearlmillet',
'head_phone']

In [ ]: