Tracking Lifecycle Behavior

Growth and Development

Raül Santaeulàlia-Llopis

MOVE-UAB and Barcelona GSE

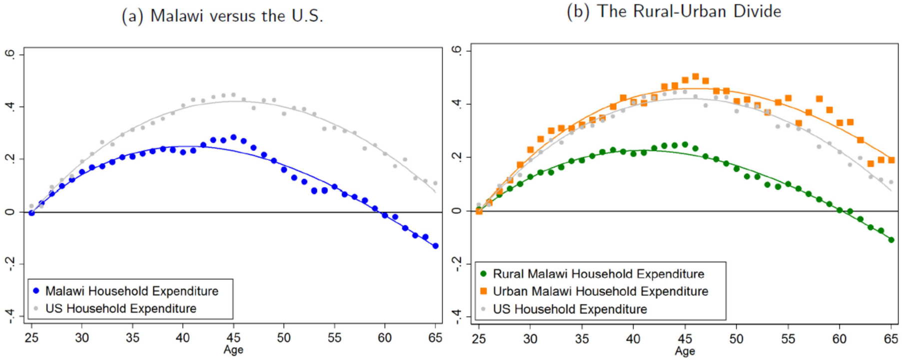
Spring 2017

Introduction

- Time, cohort and age effects: An identification issue.

- Adult-equivalent scales: A measurement (and identification) issue.

Malawi vs. the US, and the Rural-Urban Divide



De Magalhaes et al 2017.

Partial Insurance

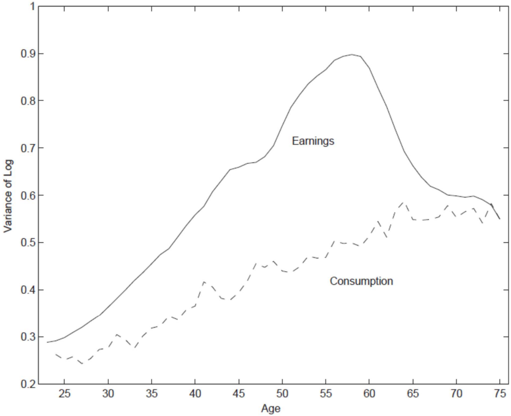


Fig. 1. The graphs represent the cross-sectional variance of the logarithm of earnings and consumption. The basic data unit is the household. Consumption data are from the CEX and are taken directly from Deaton and Paxson (1994). Earnings data are taken from the PSID. The variances are net of 'cohort effects:' dispersion which is unique to a group of households with heads born in the same year. This is accomplished, as in Deaton and Paxson (1994), via a cohort and age dummy-variable regression. The graphs are the coefficients on the age dummies, scaled so as to mimic the overall level of dispersion in the data. Further details are in Appendix A.

Measuring Lifecycle Behavior

Identifying Age, Cohort, and Time Effects

- Assume we are interested in an economic variable y for individual i that depends on his age, a, cohort[1], c and time, t

$$y_{i,a,t} = f(a, c, t, x_i, u_i)$$

  where x_i are input controls (education, family structure, wealth,...) and u_i is everything else (unobserved cross-sectional heterogeneity).

  For expositional convenience, let's drop x_i and u_i from f.

- There is no need to if our data is rich enough for each age, but otherwise we typically would group individuals in age intervals, for instance, $a - 2$ and $a + 2$.

[1] Year at birth.

- For any especification of f, note that we can always find functions

$$
\begin{aligned}
g_1(a,c) &= f(a,c,a+c) \\
g_2(c,t) &= f(t-c,c,t) \\
g_3(a,t) &= f(a,t-a,t)
\end{aligned}
$$

  such that $g_1(a,c) = g_2(c,t) = g_3(a,t)$.

- That is, we cannot identify $f(a,c,t)$ solely from the data.

- In other words, we cannot separately identify age, cohort and time effects without imposing further assumptions.

- Remark: Even with the best panel data set we face this problem.

- One especification for f is

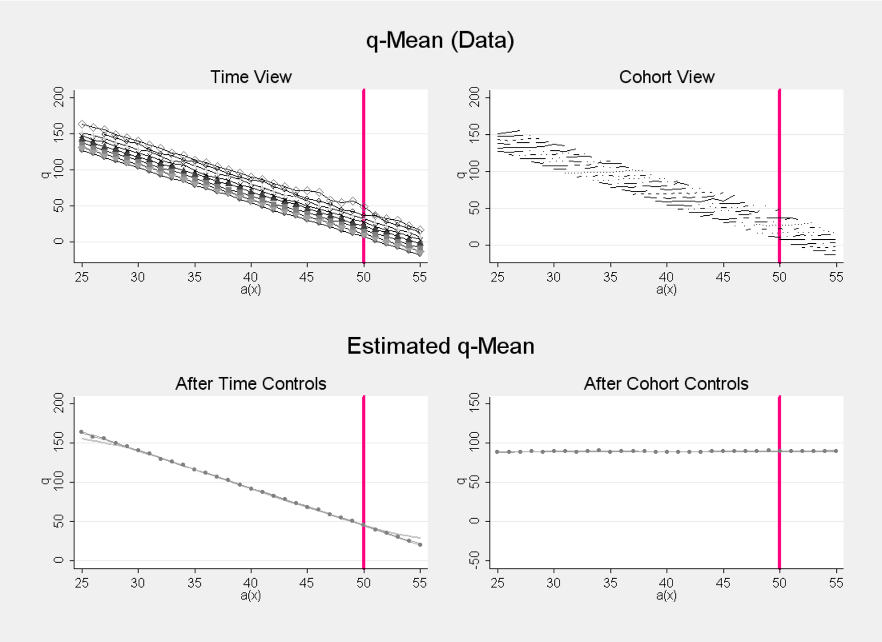
$$y_{act} = \beta_a d_a + \beta_c d_c + \beta_t d_t + \varepsilon_{bt} \tag{1}$$

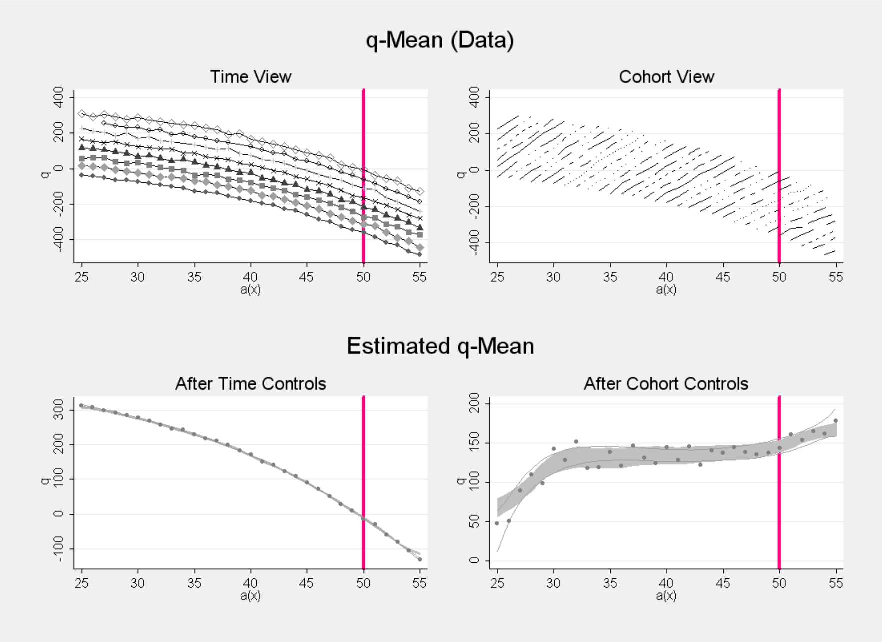
  where the d denotes dummies.

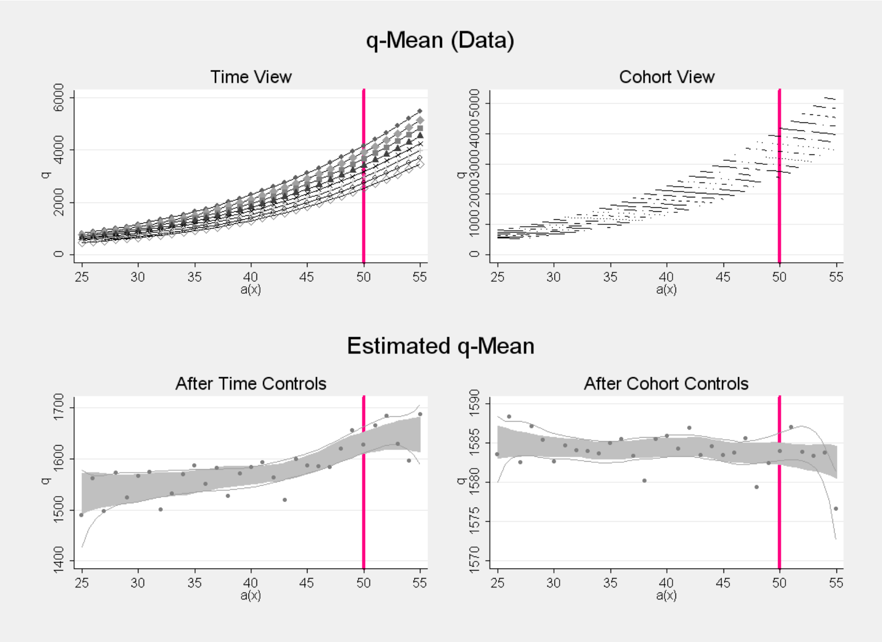
- The identification problem is that age, cohort and time are linearly dependent:

$$t = a + c$$

That is, we cannot identify the effect of a, c and $a+c$ in (1) and any trends in the data can be arbitrarily attributed to either the year effect of to a combination of age and cohort effects. We cannot empirically refute that data generated by any given $f(a, c, t)$


q-Mean (Data)
Time View
Cohort View
Estimated q-Mean
After Time Controls
After Cohort Controls
q
a(x)
25
30
35
40
45
50
55
0
50
100
150
200
-50


q-Mean (Data)
Time View
Cohort View
Estimated q-Mean
After Time Controls
After Cohort Controls
q
a(x)
25 30 35 40 45 50 55
400 200 0 -200 -400
300 200 100 0 -100
200 150 100 50 0



q-Mean (Data)
Time View
Cohort View
Estimated q-Mean
After Time Controls
After Cohort Controls
q
a(x)

More examples:

- Working example: Ameriks and Zeldes (2004)
    - Example of artificial data 1: (a) individuals of all ages hold the same c or (b) across time individuals of the same cohort increase c by 1%.
    - Example of artificial data 2: (a) c is lower for older individuals and c is raising over time conditional on age or (b) c is the same for all individuals born in the same cohort and individuals bornt at $b+1$ have 1% more of c than those born at b.
    - Example of artificial data 3: (a) the age effect is changing over time (w/o cohort effect), (b) later born cohorts increase c as they age at a faster rate than previous cohorts (w/o time effects), or (c) time effect has raised the profiles of more recently born cohorts (w/o age effect).

- Working example: Heathcote, Storesletten and Violante (2005)

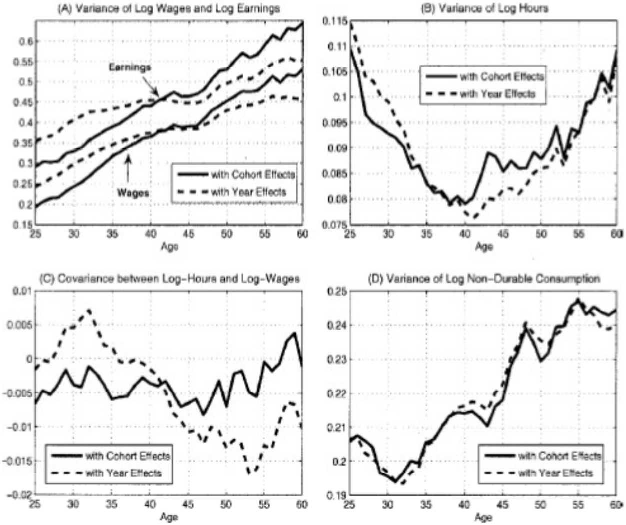


FIGURE 1. Age profiles of inequality: Cohort vs. time effects. Note: The figure portrays the 36 age dummies from running linear regressions abstracting from cohort effects and time effects, respectively. Each graph is normalized to match the unconditional average for the corresponding cross-sectional moment over the period 1980–1997.

Source: Heathcote, Storesletten, and Violante (2005)

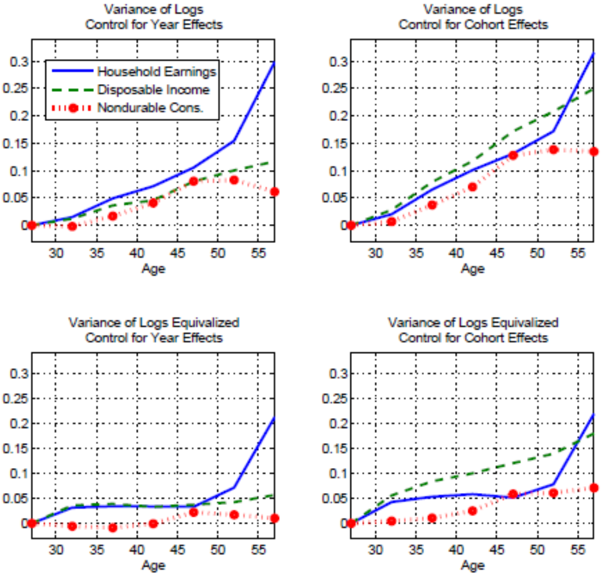


Figure 14: Life-cycle inequality: controlling for time and cohort effects (CEX)

Source: Heathcote, Perri, and Violante (2010)

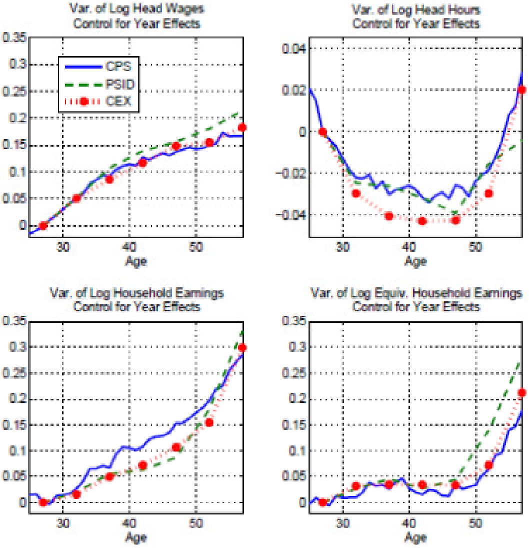
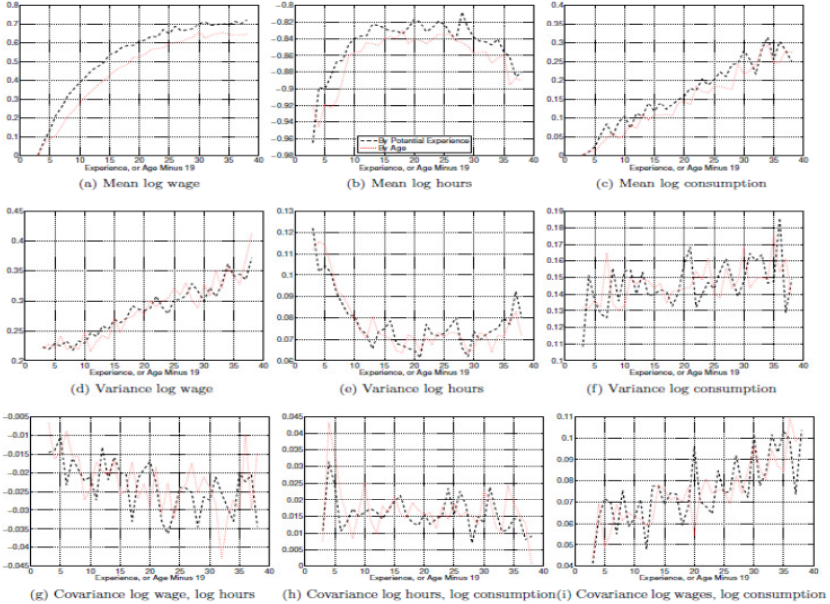


Figure 15: Comparing life-cycle inequality across datasets

Source: Heathcote, Perri, and Violante (2010)



Source: Kaplan (2010)

- Working example: Deaton (1997).

  He assumes that the time effect captures cyclical fluctuations, i.e., aggregate shocks, that average zero in the long-run and all trends can be interpreted as a combination of age and cohort effects. This implies dropping one column from the age matrix and cohort matrix of dummy variables and the first two year dummy variables, as well as normalizing the year dummy variable:

$$D^{a+b} = d_t - (t-1)d_2 + (t-2)d_1 \tag{2}$$

  where d_t is a dummy variable.

- This implies that all the year dummy variables sum to zero and makes the year effects orthogonal to a time trend:

$$\sum_{a+b=1}^{T} D^{a+b} = 0$$

$$\sum_{a+b=1}^{T} (a+b) D^{a+b} = 0$$

The coefficients of the first 2 years dummies can be then recovered given the fact that the year effects sum to zero.

- Working example: Fernández-Villaverde and Krüger (2007).

  The authors consider a seminonparametric especification when estimating consumption age profiles,

$$c_{bt} = \beta_b d_b + \beta_t d_t + m(a_{bt}) + \varepsilon_{bt}$$

  where c_{bt} is the average of log-consumption of cohort b at time t, d_b is a cohort dummy for each cohort (we have to drop one), and d_t is a dummy for each quarter (we have to drop one), a_{bt} is the age of cohort b at time t, $m(a_{bt}) = E(c_{bt}|age_{bt})$ is a smooth function of age, and ϵ_{it} is independent and with zero mean and captures measurement error (multiplicatively, because c_{bt} is in logs) and unobserved cross-sectional heterogeneity.

  The parametric bit are the dummies, the nonparametric bit is $m(a_{bt})$.

- They estimate the partially linear model with Speckman (1988) 2-step estimator that combines OLS for the parametric component and a standard kernel smooting estimator—Nadaraya-Watson estimator with Epanechnikov kernel—for the nonparametric component.

- None of this resolve the lack of identification for age, cohort and time effects. They assume that time effects are orthogonal to a time trend and that their sum is normalized to zero (as in Deaton (1997)).

Measuring Adult-Equivalent Consumption

- In the studies on consumption inequality, the researcher often only has access to household-level expenditure data.
- As welfare is typically defined for an individual (instead of a household), the common practice is to convert household-level consumption into individual-level consumption using equivalence scales.
- In particular, given that we focus on private food consumption, we usually study linear equivalence scales.

- Traditionally, the linear scale $\theta_{g,a}$ captures the inverse of the ratio between the consumption of a reference member in the household and the consumption of some non-reference household member of gender g and age a.[2]
- This way, given household consumption C,

$$C = c^h + \sum_{i \neq h} c^i,$$

we can approximate the consumption of the adult reference member as

$$c^h \approx \frac{C}{1 + \sum_{i \neq h}^{I} \mathbf{1}_{i \in \{g,a\}} \theta_{g,a}}. \quad (3)$$

We refer to c^h as the adult-equivalent consumption. Note that only information on household-level consumption, C, and on the household composition are needed to define adult-equivalent consumption, c^h.

[2] Considering the household head as the reference, the equivalence scale for a non-head individual is defined by the approximation $\frac{c^i_{g,a}}{c^h} \approx \theta_{a,g}$, where c^h is the consumption of the household head and $c^i_{g,a}$ is the consumption of the non-head household member i of gender g and age a.

- Aguiar and Hurst (2013) regress the logged household-level consumption measure on dummies for the number of adults and the number of children of different genders and ages:[3]

$$\ln C = cons + \beta_{adults} \mathbf{1}_{\#adults} + \sum_{i \neq adult}^{I} \mathbf{1}_{i \in \{g,a\}} \beta_{g,a} \tag{4}$$

The regression is run separately by area of residence (i.e. rural or urban) and by wave. Then we use the exponentiated predicted value of the regression, normalized by the value for singleton households (i.e. the exponentiated constant in the above regression), as the equivalence scale.

[3] The dummies $\mathbf{1}_{i \in \{g,a\}}$ include the dummies for the number of boys between 0 and 2, the number of girls between 0 and 2, the number of boys between 3 and 5, the number of girls between 3 and 5, the number of boys between 6 and 13, the number of girls between 6 and 13, the number of boys between 14 and 17, and the number of girls btween 14 and 17.

TABLE 1.—DIFFERENT HOUSEHOLD EQUIVALENCE SCALES

Family Size	OECD	NAS	HHS	DOC	LM	Nelson	Mean
1	1	1	1	1	1	1	1
2	1.70	1.62	1.34	1.28	1.06	1.06	1.34
3	2.20	2.00	1.68	1.57	1.28	1.17	1.65
4	2.70	2.36	2.02	2.01	1.47	1.24	1.97
5	3.20	2.69	2.37	2.37	1.69	1.30	2.27

Source: Fernandez-Villaverde and Krueger (2007)

Table 3: Parameter Estimates and Residuals of Alternative Models

	Benchmark	No Habit	Sym Habit	Sym HP	Eq Weight	OECD
θ	0.330 (0.123)	0.000 (0.102)	0.000 (0.963)	0.291 (0.187)	0.084 (0.010)	*0.7*
θ_c	2.502 (0.685)	2.272 (0.385)	2.415 (1.389)	2.565 (0.524)	3.921 (0.084)	*0.5*
θ_a	*0*	*0*	*0*	*0*	*0*	*0.7*

Source: Hong and Rios-Rull (2012)

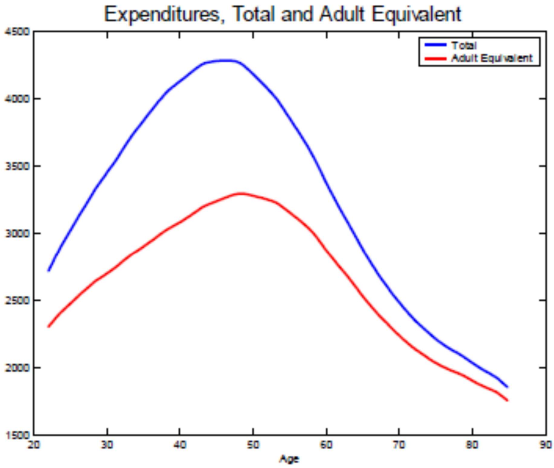


Figure 2.6: Consumption over the Life Cycle

Fernández-Villaverde and Krueger (RESTAT)

FIGURE 7.—TOTAL EXPENDITURE: ADULT EQUIVALENT, BY EDUCATION GROUPS

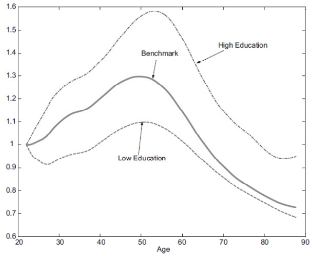


FIGURE 8.—95% CONFIDENCE BAND

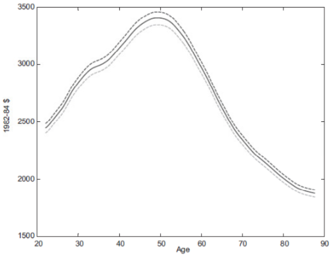
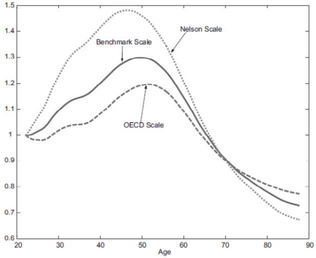


FIGURE 10.—COMPARISON OF DIFFERENT EQUIVALENCE SCALES I



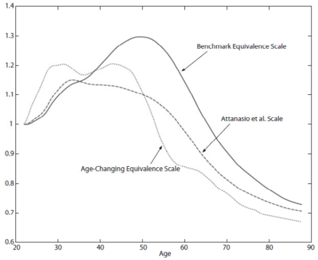
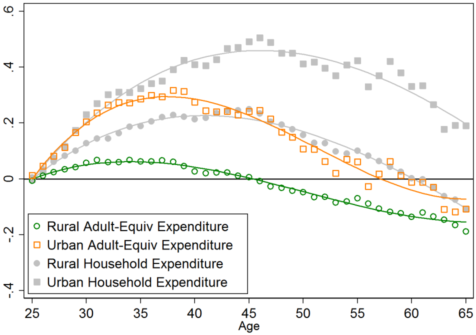


Figure: The Role of Household Structure (Malawi)



De Magalhaes et al 2017.

Figure: Lifecycle Food and Nonfood Expenditure

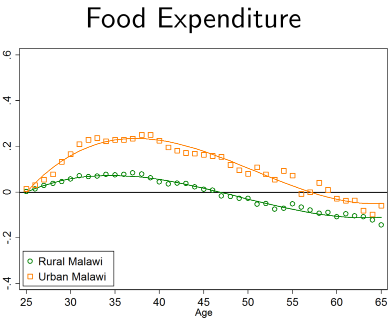
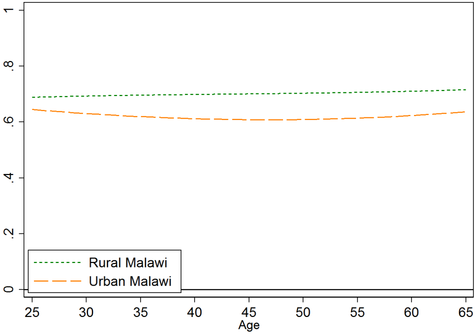
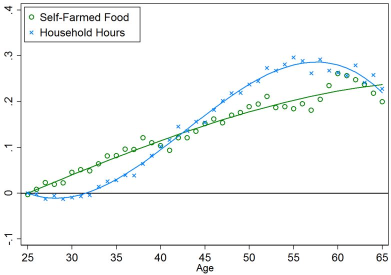


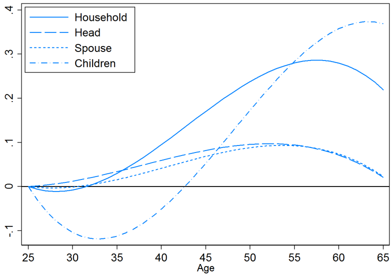
Figure: Lifecycle Shares of Food Expenditure



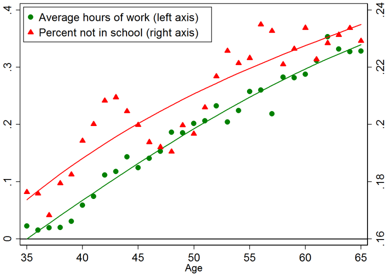
Self-Farming Food and Hours



Self-Farming Hours by Household Member



Self-Farming Hours and Schooling of School-Age Children



Are Adult-Equivalent Measures of Consumption Doing a Good Job?

Individual data is likely to help answer this question.

Table: Missing Consumption Inequality

	Rural	Urban
(a) Food Consumption:		
Individual Inequality	0.744	0.572
Adult-Equivalent Inequality	0.436	0.355
Missing Inequality (%)	41.48	37.71
(b) Core Food Consumption (Excl. "Vices"):		
Individual Inequality	0.311	0.275
Adult-Equivalent Inequality	0.252	0.234
Missing Inequality (%)	19.26	14.76

Notes: Missing consumption inequality is defined as the share of actual individual consumption inequality not captured by adult-equivalent consumption inequality, that is, $\sum_t 100 \times \left(1 - \frac{var_t(\ln c^h)}{var_t(\ln c^i)}\right)$, where $var_t(\ln c^h)$ is the cross-sectional variance of logged adult-equivalent consumption and $var_t(\ln c^i)$ is the cross-setctional variance of actual individual consumption. We report the average across our eight waves of CHNS data, 1991-2011.
Authors: Santaeulàlia-Llopis and Zheng (2016)

Table: Missing Consumption Inequality: Households without Children

	Rural	Urban
(a) Food Consumption:		
Individual Inequality	0.727	0.559
Adult-Equivalent Inequality	0.512	0.408
Missing Inequality (%)	29.65	26.96
(b) Core Food Consumption (Excl. "Vices"):		
Individual Inequality	0.282	0.257
Adult-Equivalent Inequality	0.267	0.245
Missing Inequality (%)	5.32	4.70

Notes: Missing consumption inequality is defined as the share of actual individual consumption inequality not captured by adult-equivalent consumption inequality, that is, $\sum_t 100 \times \left(1 - \frac{var_t(\ln c^h)}{var_t(\ln c^i)}\right)$, where $var_t(\ln c^h)$ is the cross-sectional variance of logged adult-equivalent consumption and $var_t(\ln c^i)$ is the cross-setctional variance of actual individual consumption. We report the average across our eight waves of CHNS data, 1991-2011. Authors: Santaeulàlia-Llopis and Zheng (2016).