Supplementary Information

Asset index

The purpose of an Asset Index (AI) is to develop a proxy measure for a household's socio-economic position or long-run socio-economic status (Filmer & Pritchett 2001). Direct questions concerning wealth were not included in the questionnaire owing to their noted unreliability (due to false reporting or recall bias) and short-term variability (due to seasonality for example). An AI uses proxy measures (such as the physical attributes of the house or social characteristics of the household members) to generate insights around financial stock (permanent income) rather than flow (current income) (Balen et al. 2010). Principle Components Analysis (PCA) provides a means to evaluate the most meaningful aspects of the large amounts of data that are often generated about the physical and social attributes of the survey respondents, thus revealing the underlying structure of the data. The underlying data structure is then used to generate a single measure for each household that represents a households long-run socio-economic position. This measure can be used to compare different households.

Table SI1 shows the list of variables, which were initially considered as potential elements of an AI. The list of variables was identified from existing theory concerning asset indexes and insights gained through the field research and subsequent analysis (see, for example, Filmer & Pritchett 2001; Balen et al. 2010; Hunter et al. 2014; You 2014).

Table SI1: The initial variables from which the asset index was potentially constructed.

Variable	Description
Bdrms	Number of bedrooms in property
Heating	Presence of artificial heating (charcoal, wood, electricity)
Ceiling_fan	Number of ceiling fans in the property
Refrigerator	Number of refrigerators
Washing_machine	Number of washing machines
Bicycle	Number of bicycles
M.bike_scooter	Number of motorbikes or three-wheeled scooters
Savings	Presence of savings
Loans	Presence of a loan
HHsize	Number of household members (adults and children)
HHAdult	Number of adults in the household
HHChild	Number of children in the household
Wellbeing	Overall wellbeing of the household
Old.dependants	Number of elderly dependents
Young.dependants	Number of young dependents
Dependants	Overall number of dependents
Dependancy.ratio	The ratio of the number of dependents in the household to the number of economically active adults

Table SI1 (cont.): The initial variables from which the asset index was potentially constructed.

Variable	Description
TotalMU	Total area of farmland available to the household
PlotNum1	Number of plots
AvPlot1	Average plot size
Agric_dependency	Perceived dependency on agriculture
Irrigation	Availability of irrigation
Percent_irrig	Percentage of farmland household able to irrigate
Livestock	Number of livestock owned
Chickens	Number of chickens owned
Fowl	Number of fowl owned
Goats	Number of goats owned
Pigs	Number of pigs owned
OtherFarm	Participation in other farm-related activities (such as fishing or forestry)
Fishing	Participation in fishing
Forestry	Participation in forestry-related activities
Age	Age of household head
School	Number of years the household head attended school
Health	Health status of household head
Remittances	Receipt of remittances

To generate the AI a number of steps were required to transform the data. All non-numeric variables were transformed into numeric values. For discrete variables, where the relative difference between sources was indeterminable, the categories were converted to a simple dichotomous variable (1 = no and 2 = yes) to indicate the presence or absence. For example, the absence of a power supply was coded as 1 and all other power sources (charcoal, electricity, wood and other) were coded as 2. All other discrete variables were recoded with higher numbers signifying increasingly positive attributes. All dichotomous variables recoded to 1 = no and 2 = yes. Don't know responses were interpolated to the mean sample value for that variable. The working sample of households (n=73) consisted of only those cases that were complete.

The transformed data was organised into a correlation matrix to check for internal consistency and uncorrelated and multicollinear variables, variables were removed with a large number of correlations that were <0.2 or >0.9. Factorability was determined through Bartlett's test of sphericity[1] and the Kaiser-Meyer-Olkin (KMO) test[2]. Variables were excluded (based on Barlett's test for spericity and the KMO test) in a stepwise fashion to identify a correlation matrix suitable for PCA. Once a suitable group of variables were identified, a further diagnostic tool was used: the

[1] Bartlett's test of sphericity helps to reveal if the correlation matrix is an identity matrix (all variables are completely independent of each other) and not suitable for principle components analysis. An identify matrix is suggested if Bartlett's test was not significant (value of >0.05) (Field et al. 2012).

[2] The KMO test reveals the diffusion in the pattern of correlations. A value close to 1 suggests that the data are relatively compact and that factor analysis would provide reliable distinct factors. Values close to 0 suggest that there is diffusion in the pattern of correlations and that a factor analysis will be inappropriate. For the correlation matrix to be considered acceptable a value ≥0.7 was required. The KMO test also provides values for individual variables; variables retained with higher scores, values <0.7 sought for the majority of variables (Field et al. 2012).

determinant of the correlation matrix[3] (Field et al. 2012). An eight variable correlation matrix was identified as suitable for PCA based on the diagnostic tests outlined above.

PCA was run on all variables to reduce the data down to its underlying factors. The number of factors to extract was determined by visual inspection of scree plots, cumulative proportion of variance and Joliffe's criterion[4] . Based on the aforementioned tests, all eight factors were retained and the standardised AI computed using principle (first) component factor loadings or weights (Field et al. 2012). Table SI2 shows the variables that were used in the PCA with some summary statistics and the weighting derived from the principle (first) component analysis. The final AI was derived from the PCA weights and transformed to include only positive numbers and a range from zero to one.

Table SI2: Variables used in the PCA with some summary statistics and weighting derived from the principle (first) component analysis (all values to 2dp).

Variable	Mean	S.D	Variance	Weight
Ceiling_fan	2.26	1.18	1.39	0.94
Refrigerator	0.62	0.52	0.27	0.12
Washing_machine	0.64	0.48	0.23	0.27
Bicycle	0.74	0.78	0.61	0.08
HHAdult	2.33	0.76	0.58	0.12
Dependency.ratio	0.24	0.25	0.06	0.08
TotalMU	9.06	4.05	16.42	0.08
Age	60.23	12.82	164.32	-0.06

Having calculated the AI for the working sample the process was repeated on the total sample (n=97). Missing values for each variable were populated through interpolation (values for interpolation were derived by calculating the mean of each variable for either case study site).

Table SI3 shows the AI for each case study site and both sites together. The table shows that Dongdian has a higher mean AI score compared to Wanzhuang suggesting that the households are in a slightly better socio-economic position.

Table SI3: The values of the AI for both sites and each site individually.

Site	Sample size	AI (4dp)
Dongdian	50	0.4913
Wanzhuang	47	0.3973
Both sites	97	0.4458

A typical household in the top quartile of the AI would tend to come from Dongdian (by a ratio of 2: 1) and have three to four bedrooms with some source of artificial heating. Most households own a

[3] The determinant of the correlation matrix shows the extent to which the data are singular (value = 0) or unrelated (value =1). A value >0.00001 is necessary for a PCA (Field et al. 2012).
[4] Joliffe's criterion suggests that factors with eigenvalues ≥0.7 are retained.

washing machine and a refrigerator and all have a motorcycle or three-wheeled scooter to get around in addition to a bicycle. There tend to be between three and four adults per household (average age of about 50 years) with a low dependency ratio (0.33). Most households class themselves dependent on agriculture with larger than average farm sizes. Irrigation is only available to approximately half of this group and only 30 per cent of farmland is available for irrigation within this subset. Most households practice some sort of other farm activities (small-scale animal husbandry) in addition to crop growing.

A typical household in the bottom quartile of the AI is more likely to come from Wanzhuang (by a ratio of 1: 1.4) although the split is more even when compared to the top quartile. Houses tend to be slightly smaller with fewer bedrooms and living spaces compared to the top quartile households. These households are much less likely to own a washing machine or a refrigerator and less than half have access to a motorcycle, three-wheeled scooter, or a bicycle. There tend to be around two adults per household (average age of about 70 years). Interestingly, the dependency ratio (0.15) for this group of households is lower than the upper quartile households. As with the upper quartile most class themselves as dependent on agriculture although farm sizes are much smaller and irrigation is only available to a third of this group (although a comparable area is irrigated). The households do practice other non-crop activities in regard to farming but there is less diversity and it tends to be on a smaller scale.

Indices capturing self-perceived measures of change

The purpose of constructing indices for perceived changes to yield, finance, wellbeing, and agricultural dependency was to gain a comparable insight into how different groups of households perceived change and the direction of that change over time. For example, through the indices one could compare the perceptions of households with migrant members and households with no members practising mobility. The indices were all constructed using the same simple methodology. Information on changes to yield, finance, wellbeing, and agricultural dependency (at date of survey compared to five years ago) were collected through the household survey. Respondents were asked to respond to four questions in a closed format outlined in table SI4 below.

Table SI4: Variable and list of options possible for respondents to select.

Variable	Yield	Finance	Wellbeing	Agricultural dependency
Question in household survey	Compared to 5 years ago have your crop yields changed?	Compared to 5 years ago do you think your household's financial position has changed?	Compared to 5 years ago do you think your household's wellbeing has changed?	How dependent are you and your household on agriculture for your livelihoods?
Possible responses	Increased	Got better	Got better	Completely dependent
	Decreased	Got worse	Got worse	Dependent
	Stayed the same	Stayed the same	Stayed the same	Not that dependent
	Not comparable			Not dependent at all

Responses from the household survey were filtered to exclude those who stated that they were unable to recall the flood and drought events or were incomplete (precluding subsequent analysis). Additionally, one respondent stated that he or she was unable to make a meaningful comparison for crop yield and was also excluded. The data processing and cleaning resulted in a useable sample of 62 households, of which 33 were from Dongdian and 29 were from Wanzhuang (see Table SI5 and Table SI6).

Table SI5: Responses to questions capturing changes in yield, finance, and wellbeing (at date of survey compared to five years ago) (n=62).

	Positive change	No change	Negative change
Yield	39	12	11
Finance	36	3	23
Wellbeing	46	4	12

Table SI6: Responses to questions capturing perceived dependency on agriculture for household livelihood changes (at date of survey compared to five years ago) (n=62).

	Completely dependent	Dependent	Not that dependent	Not dependent
Agricultural dependency	24	23	11	4

Some simple common transformations were applied to the data to create a single value for different groups of households. For yield, finance, and wellbeing, the positive change value was multiplied by three, the no change value was multiplied by two, and the negative change value was multiplied by one. The totals were summed and divided by three to give one value for each group of respondents and converted to per cent. A similar process was repeated for agricultural dependency although, as there were four categories, the completely dependent value was multiplied by four with the subsequent categories multiplied by three, two and one respectively. The total was summed, as with yield, finance and wellbeing transformation, divided by four (as there were four categories rather than three) and converted to per cent. Table SI7 shows some simple summary statistics for the entire sample (n=62).

Table SI7: Simple summary statistics of the transformed data capturing changes in yield, finance and wellbeing and agricultural dependency (compared to five years ago) (n=62; all values to 2dp).

	Mean	S.D	Variance
Yield	82.75	7.04	49.58
Finance	75.25	4.50	20.25
Wellbeing	85.00	6.73	45.33
Agricultural dependency	75.00	7.62	58.00

Index capturing externally-perceived level of wellbeing

The index representing externally perceived measure of wellbeing was derived from the output of a rapid rural appraisal activity (RRA). The output of the RRA activity ranked all members of the community according to the perceived level of wellbeing. The ranking provided a comparative measure showing how the participants of the RRA exercise perceived the members of the community at that point in time (it does not capture change over time). In both case study sites, the participants of the RRA activity created three groupings, representing different categories of wellbeing (upper, middle and lower). Using the same methodology outlined above, these categories were converted into a single value. The upper group was multiplied by three, the middle group was multiplied by two, and the lower group was multiplied by one. The totals were summed and divided by three to give one value for each group of respondents and converted to per cent (see Table SI8 for some simple summary statistics).

Table SI8: Simple summary statistics of the transformed data capturing externally perceived measure of wellbeing (n=62; all values to 2dp).

	Mean	S.D	Variance
Wellbeing ranking	61.75	6.60	43.58

References

Balen, J., Mcmanus, D., Li, Y.-S., Zhao, Z.-Y., Yuan, L.-P., Utzinger, J., Williams, G., Li, Y., Ren, M.-Y., Liu, Z.-C., Zhou, J. & Raso, G. 2010. Comparison of Two Approaches for Measuring Household Wealth Via an Asset-Based Index in Rural and Peri-Urban Settings of Hunan Province, China. *Emerging Themes in Epidemiology,* 7**,** 1-17.

Field, A. P., Miles, J. & Field, Z. 2012. *Discovering Statistics Using R,* London, SAGE.

Filmer, D. & Pritchett, L. H. 2001. Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. *Demography,* 38**,** 115-132.

Hunter, L. M., Nawrotzki, R., Leyk, S., Maclaurin, G. J., Twine, W., Collinson, M. & Erasmus, B. 2014. Rural Outmigration, Natural Capital, and Livelihoods in South Africa. *Population, Space and Place,* 20**,** 402-420.

You, J. 2014. Risk, under-Investment in Agricultural Assets and Dynamic Asset Poverty in Rural China. *China Economic Review,* 29**,** 27-45.