Data Description

The main data we use is the *National Fixed Point Survey* (NPFS), which is a nationally representative panel dataset (unbalanced) of roughly 20,000 households in 360 villages between 1986 and 2013. It is collected by the Ministry of Agriculture and Rural Affairs of China. Since we are learning simple linear regression and multiple linear regression at the current stage, we provide a cross sectional dataset this time which only includes households in the year of 2010.

Variable Name	Type	Variable Label
year	int	
id	str8	Household id
c10	double	Arable area subcontracted into (mu)
c13	double	Arable area subcontracted out (mu)
d31	double	Total sown area: grain crops (mu)
d32	double	Total output: grain crops (kg)
f1	double	Purchase quantity: grain crops seeds (kg)
f2	double	Purchase value: grain crops seeds (yuan)
f9	double	Purchase quantity: fertiliser (kg)
f10	double	Purchase value: fertiliser (yuan)
f25	double	Purchase quantity: agricultural diesel (kg)
f26	double	Purchase value: agricultural diesel (yuan)
f27	double	Purchase quantity: agricultural firm (kg)
f28	double	Purchase value: agricultural firm (yuan)
f29	double	Purchase quantity: agrochemicals (kg)
f30	double	Purchase value: agrochemicals (yuan)
g1	double	Total annual household income (yuan)
g44	double	Income from working outside home (labour income, yuan)
g45	double	Income from business outside the home (yuan)
g50	double	Income from interest, dividend and bonus (yuan)
g81	double	Total annual household expenditures (yuan)
g82	double	Household operating costs (yuan)
g110	double	Expenses - living & consumption (yuan)
g113	long	Expenses - consumption (yuan)
g114	double	Expenses - consumption - food (yuan)
g149	double	Annual household net income (yuan)
g150	long	Annual household net productive income (yuan)
h47	double	Residential housing at the end of the year (square metre)
huzhu_gender	float	Gender of householder (0: male, 1: female)
huzhu_age	float	Age of householder
huzhu_edu	float	Education attainment of householder
health	float	Percentage of household population in good health condition
vl_id	str4	Village id