The Economic Journal, Doi: 10.1111/j.1468-0297.2011.02452.x. © 2011 The Author(s). The Economic Journal © 2011 Royal Economic Society. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

Technical Appendix to

DID HIGHER INEQUALITY IMPEDE GROWTH IN RURAL CHINA?

Dwayne Benjamin, Loren Brandt and John Giles

Economic Journal, doi: 10.1111/j.1468-0297.2010.02452.x

Appendix S1: Relating the Household and Village-level Specifications

As our estimation is conducted at the household-level, we avoid most issues of aggregation. However, it is still worth relating our approach in (1) to the more conventional village-level specification, especially since all of the key variation (of inequality) resides at the village-level. To begin with, consider a simplified village-level specification relating growth to inequality (the Mean Log Deviation):

$$g_{vT} = \overline{\ln y}_{vT} - \overline{\ln y}_{vt-1} = \beta_0 + \beta_1 \overline{\ln y}_{vt-1} + \beta_v \left(\ln \bar{y}_{vt-1} - \overline{\ln y}_{vt-1}\right) + \varepsilon_{vT}. \tag{A.1}$$

Note that this is a departure from the previous literature as we define growth as the difference in the mean log incomes (i.e. average household growth rate), as opposed to the change in log mean incomes. This allows cleaner aggregation from the household-level. There are several ways to obtain (A.1) by averaging a household specification.

Consider first a household-level model that allows for an external effect of mean village (log) income, as well as own income, plus village income inequality:

$$g_{i,vT} = \ln y_{i,vT} - \ln y_{i,vt-1} = \alpha_0 + \alpha_1 \ln y_{i,vt-1} + \alpha_2 \overline{\ln y}_{vt-1} + \alpha_v \left(\ln \bar{y}_{vt-1} - \overline{\ln y}_{vt-1}\right) + \varepsilon_{i,vT}. \tag{A.2}$$

This averages by village to:

$$g_{vT} = \alpha_0 + (\alpha_1 + \alpha_2) \overline{\ln y}_{vt-1} + \alpha_v \left(\ln \bar{y}_{vt-1} - \overline{\ln y}_{vt-1}\right) + \varepsilon_{vT}. \tag{A.3}$$

Notice that the coefficient on $\overline{\ln y}_{vt-1}$ captures the combined effect of own (log) household income and mean (log) village income. Aggregation can also be accomplished by using village-dummies as instruments. By projecting everything to the village-level, it is clear that effects of own and village-level mean income are not separately identified at the village-level: the village-level specification combines both ($\alpha_1 + \alpha_2$). This is related to the main point raised in Ravallion (1998), though the confounding of the own-income effect with inequality is not an issue with our specification, as we use 'mean logs' as opposed to 'log means' as our key village income variables.

A second way to obtain a version of (A.1) is to specify the household-level equation as:

$$g_{i,vT} = \alpha_0 + \alpha_1 \ln y_{i,vt-1} + \alpha_2 \overline{\ln y}_{vt-1} + \alpha_v \underbrace{\left(\ln \bar{y}_{vt-1} - \ln y_{i,vt-1}\right)}_{\mathrm{DEV}_{i,vt-1}} + \varepsilon_{i,vT}. \tag{A.4}$$

The key difference between (A.2) and (A.4) is that the inequality measure is $\text{DEV}_{i,vt-1} \equiv \ln \bar{y}_{vt-1} - \ln y_{i,vt-1}$, the amount by which household i's income is less than the log of the mean income in the village (controlling for both the independent effect of own log income and the mean log income in the village). This equation aggregates to the same village-level regression (A.1). At the village-level, we cannot tell whether it is MLD or DEV that affects average household income growth. From our perspective, the distinction is not important, as there is no 'right' measure of inequality.

As noted earlier, the village-level equation (A.1) can be estimated with the household-level data and specification (A.4) by 2SLS, using as instruments either a vector of village dummies, or identically a vector of village means, $\overline{\ln y}_{vt-1}, \text{MLD}_{vt-1}$ (see, for example, Angrist and Pischke (2009) and 'Visual Instrumental Variables'). As long as the weights are correct (i.e. use the number of households in a village as weights in the village-level specification), then the household (2SLS) and village-level (WLS) specifications will yield identical coefficients for α_v, β_v.

While it is conventional to estimate the village-level (or country-level) regression using village sample averages and measures of inequality (e.g. $\overline{\ln y}_{vt-1}, \text{MLD}_{vt-1}$), there are small-sample and measurement-error issues with using sample averages as 'proxies' for population moments (Deaton, 1985). Devereux (2007) shows, however, that the Deaton measurement-error estimator is identical to a Jackknife Instrumental Variable (JIVE) estimator, where jackknifed sample means serve as instruments for individual-level variables. The intent in that exercise, it should be noted, is NOT to estimate or identify the household-level specification, but to correct for measurement error in the village-level (aggregate) specification. Adapting Devereux's framework, the implied 'structural' equation at the household-level is:

$$g_{i,vT} = \alpha_0 + (\alpha_1 + \alpha_2) \ln y_{i,vt-1} + \alpha_3 \underbrace{(\ln \bar{y}_{vt-1} - \ln y_{i,vt-1})}_{\text{DEV}_{i,vt-1}} + \varepsilon_{i,vT}. \tag{A.5}$$

Correcting for measurement error (implementing Deaton's estimator using Devereux's jackknife result) simply involves using $\overline{\ln y}_{(-i)vt-1}$ and $\text{MLD}_{(-i)vt-1}$ as instruments for $\ln y_{i,vt-1}$ and $\text{DEV}_{i,vt-1}$. The implied reduced form equation for $g_{i,vT}$ in this approach is:

$$g_{i,vT} = \pi_0 + \pi_1 \overline{\ln y}_{(-i)vt-1} + \pi_2 \text{MLD}_{(-i)vt-1} + v_{i,vT} \tag{A.6}$$

Except for our inclusion of covariates (which can be easily incorporated into Devereux's framework), this reduced form (A.6) is identical to our structural equation (1). In other words, our main estimating equation can be seen as 'highly similar' to the implied reduced form associated with Deaton's measurement-error correcting group-means estimator.[1] It is not necessary that our model be imbedded within this framework, as (1) is a *bona fide* structural model in its own right. However, (A.6) allows us to illustrate better the advantages of using the household-level data to correct for a variety of potential econometric problems associated with using the aggregated data and to highlight the role of aggregation precisely.

[1] There is one other bit of slippage: in this discussion we treat $\ln \bar{y}_{vt-1}$ as a known parameter. Of course, it is not. However, because the sample mean is inside the non-linear logarithm function, the linear algebra associated with the JIVE estimator does not hold exactly (without treating it as a constant). In our empirical work, however, we use the jackknifed version of $\bar{y}_{(-i)vt-1}$ inside the logarithm to address 'in spirit' the same small-sample (measurement error) problem. Note that this only effects the degree to which our equation (1) is an *exact* reduced form for the Deaton/Devereux specification.

Table S1

Attrition Equations Probit Estimates: Probability of Being in the Balanced Panel Estimating Sample

	In panel sample	
	(1)	(2)
Initial village inequality (Mean Log Deviation)	−0.516	−0.920
	(1.698)	(1.726)
Household variables (in initial period, 1987–8)		
Log initial income	18.614*	−0.195*
	(5.633)	(0.063)
Log initial income – squared	−2.873*	
	(0.908)	
Log initial income – cubed	0.144*	
	(0.049)	
Share of income from agriculture in 1987–8	0.180	0.197
	(0.217)	(0.217)
Share of income from wages in 1987–8	0.082	0.093
	(0.192)	(0.196)
Share of income from family businesses in 1987–8	0.010	−0.012
	(0.185)	(0.184)
Household education in 1987–8	−0.004	−0.003
	(0.010)	(0.010)
Log household size	0.087	0.097
	(0.074)	(0.073)
Dependency ratio	0.183	0.190
	(0.143)	(0.142)
Log cultivated land	0.009	0.010
	(0.023)	(0.024)
Head age $\leq$ 30	−0.089	−0.094
	(0.078)	(0.078)
Head age between 31 and 40	0.030	0.029
	(0.052)	(0.052)
Head age between 51 and 60	−0.184*	−0.180*
	(0.065)	(0.064)
Head age 61 and over	−0.275*	−0.257*
	(0.098)	(0.097)
Village variables (in initial period, 1987–8)		
Mean log per capita income	−0.095	−0.074
	(0.256)	(0.259)
Average education	0.049	0.054
	(0.065)	(0.065)
Cultivated land per capita	−0.115	−0.093
	(0.153)	(0.155)
Village dependency ratio	0.193	0.369
	(1.598)	(1.583)
Village share of income from agriculture	−0.652	−0.563
	(1.492)	(1.458)
Village share of income from wages	−0.650	−0.504
	(1.522)	(1.500)
Village share of income from family businesses	−0.677	−0.462
	(1.700)	(1.642)
Village tax revenue per capita	0.954*	0.847*
	(0.427)	(0.432)

Table S1
(*Continued*)

	In panel sample	
	(1)	(2)
Village government expenditure per capita	−0.846*	−0.769
	(0.405)	(0.411)
Sample size (1987–8)	4,847	4,847

Notes. The reported coefficients are from a Probit where the dependent variable is the probability of being in the primary estimating sample (i.e. the balanced panel observed in all years from 1987 through 2002, with complete variable set). Robust, cluster-corrected standard errors are reported in parentheses, with * indicating statistical significance at the 5% level. All specifications include village location and province dummies. Specification (1), with the cubic in household income, is used to generate inverse probability weights for the attrition corrections throughout the paper.

Table S2
Additional Covariates (Corresponding to Table 2, Column 5)

Endpoint	1990–1	1995–6	1997–8	1999–2000	2001–2
Initial inequality (MlnD, jackknifed)	−0.354*	−0.240*	−0.197*	−0.173*	−0.111
	(0.173)	(0.084)	(0.085)	(0.071)	(0.068)
Household variables					
ln initial income (Y0)	−0.321	−0.133	−0.249	−0.024	−0.252
	(0.938)	(0.234)	(0.278)	(0.210)	(0.185)
(ln Y0) – squared	0.014	0.009	0.028	−0.009	0.027
	(0.149)	(0.039)	(0.046)	(0.034)	(0.030)
(ln Y0) – cubed	0.000	−0.001	−0.002	0.001	−0.001
	(0.008)	(0.002)	(0.003)	(0.002)	(0.002)
Household education 1987–8	0.004*	0.002*	0.000	0.001	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Log family size	0.013	0.011*	0.008*	0.005	0.000
	(0.009)	(0.003)	(0.003)	(0.002)	(0.002)>
Dependency ratio	−0.041*	0.013	0.024*	0.029*	0.030*
	(0.014)	(0.007)	(0.006)	(0.005)	(0.005)
Log cultivated land	0.000	0.001	0.000	−0.002*	0.001
	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
Age $\leq$ 30	−0.002	−0.011*	−0.008*	−0.006	−0.005
	(0.009)	(0.004)	(0.003)	(0.003)	(0.003)
31 $\leq$ Age $\leq$ 40	−0.024*	−0.011*	−0.007*	−0.004*	−0.004*
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)
51 $\leq$ Age $\leq$ 60	0.002	−0.006	−0.009*	−0.007*	−0.010*
	(0.007)	(0.003)	(0.002)	(0.002)	(0.002)
Age > 60	−0.035*	−0.027*	−0.018*	−0.016*	−0.019*
	(0.009)	(0.005)	(0.004)	(0.003)	(0.003)
Village variables					
Log mean initial income	0.023	0.026	0.034*	0.028*	0.017
	(0.028)	(0.014)	(0.014)	(0.011)	(0.009)
Avg. education	−0.001	0.001	0.002	0.003	0.003
	(0.006)	(0.003)	(0.003)	(0.002)	(0.002)
Log per capita land	0.033*	0.006	0.006	0.003	0.003
	(0.015)	(0.008)	(0.008)	(0.007)	(0.006)

Table S2
(*Continued*)

Endpoint	1990–1	1995–6	1997–8	1999–2000	2001–2
Dependency ratio	−0.336*	−0.109	−0.080	−0.120*	−0.116*
	(0.144)	(0.066)	(0.057)	(0.044)	(0.037)
Share of HH Income from agriculture	−0.063	0.035	0.015	0.058	0.045
	(0.143)	(0.052)	(0.053)	(0.049)	(0.042)
Share of HH income from wages	0.047	0.068	0.019	0.057	0.024
	(0.160)	(0.054)	(0.053)	(0.050)	(0.046)
Share of HH income from family business	−0.004	0.059	0.020	0.093	0.045
	(0.150)	(0.067)	(0.067)	(0.057)	(0.051)
Tax revenue per capita	−0.028	−0.001	−0.011	−0.005	−0.006
	(0.056)	(0.022)	(0.020)	(0.015)	(0.015)
Gov. expenditure per capita	0.050	0.003	0.011	0.010	0.015
	(0.061)	(0.021)	(0.020)	(0.016)	(0.017)
N	3,424	3,424	3,424	3,424	3,424

Notes. This table reports coefficients (with standard errors in parentheses) of additional covariates from the specification reported in column (5) of Table 2. All specifications employ attrition weights. Village mean log income and education are jackknifed. For further details, see the notes to Table 2.