Lecture 2: Barriers to Schooling

A. Low levels of health impede learning, role of gender.

Health interventions: Kenya de-worming RCT, Bangladesh clean water initiatives.

B. Burden of large families on schooling investment.

Population control policies, including direct control of fertility (China "One Child" policy).

C. High opportunity costs of schooling for low-income households.

Mexican *progresa* reduced these costs (RCT analysis).

The Effect of Health on Schooling

Poor health, related to bad endowments - geography - often thought to be one reason for lack of growth.

One linkage - poor health among children reduces school attendance, performance.

> 1.3 billion people infected by hookworm and roundworm; whipworm affects 900 million; schistosomiasis affects 200 million. Children account for bulk of infections - 85-90% of all heavy schistosomiasis infections in Eastern Kenya.

De-worming is relatively cheap - single-dose oral treatment reduces infections by 99%, but need annual application because of reinfection.

Prior studies, mostly RCT's, find little effect of de-worming treatment on school performance - test scores.

Randomization of treatment of children *within* schools.

Problem: This design neglects externalities: non-treated children are less likely to be infected.

Understates the effect of treatment for two reasons:

1. Control group actually affected by the treatment: difference between control-group and treatment-group outcomes underestimates effectiveness of treatment on the treated.

2. Ignores the reduction in the infections of the untreated - the externality - which is part of the benefits of the treatment.

Miguel and Kremer (2003) de-worming RCT is thus an important study.

Methodology:

A. Randomized phase-in of treatment at the school level, not individual.

    Location; Busia, Kenya.

B. Identifies the externality benefits.

Also, long-term follow-ups:

Able to quantify the effect of a two-year de-worming (only) in primary school on adult outcomes.

Estimation strategy to take into account externalities:

Takes advantage of the fact that many close neighbors go to different schools: people in proximity get different treatments

Children attending school or living nearby treatment schools have different exposure to risk of infection: exposure is number of pupils in treatment school that are nearby, or distance times treatment pupil density

Thus estimating equation is:

$$Y_{ijt} = a + \beta_1 T_{1it} + \beta_2 T_{2it} + \Sigma_d(\gamma_d N^T_{dit}) + \Sigma_d(\gamma_d N_{dit}) + u_i + e_{ijt}$$

where Y_{ijt} = outcome for student j in school i at time t
T = assigned treatment in year 1 or 2
N_{dit} = total number of pupils in primary schools at distance d from school i
N^T_{dit} = number of pupils in treatment schools at distance d from school i
(d = 1 is 1 kilometer, d=2 is 2 kilometers, etc.)
u_i = school effect

So, average treatment effect is $\beta_{1t} + \Sigma_d(\gamma_d N^{T'}_{dit})$, where $N^{T'}$ is the average number of pupils in treatment schools located at d from the school

Results

Table 9: School participation, direct effects and externalities†
Dependent variable: Average individual school participation, by year

	OLS (1)	OLS (2)	OLS (3)	OLS (4) May 98-March 99	OLS (5) May 98-March 99	OLS (6) May 98-March 99	IV-2SLS (7) May 98-March 99
Moderate-heavy infection, early 1999						-0.028*** (0.010)	-0.203* (0.094)
Treatment school (T)	0.051*** (0.022)						
First year as treatment school (T1)		0.062*** (0.015)	0.060*** (0.015)	0.062* (0.022)	0.056*** (0.020)		
Second year as treatment school (T2)		0.040* (0.021)	0.034* (0.021)				
Treatment school pupils within 3 km (per 1000 pupils)			0.044** (0.022)		0.023 (0.036)		
Treatment school pupils within 3-6 km (per 1000 pupils)			-0.014 (0.015)		-0.041 (0.027)		
Total pupils within 3 km (per 1000 pupils)			-0.033** (0.013)		-0.035* (0.019)	0.018 (0.021)	0.021 (0.019)
Total pupils within 3-6 km (per 1000 pupils)			-0.010 (0.012)		0.022 (0.027)	-0.010 (0.012)	-0.021 (0.015)
Indicator received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)					0.100*** (0.014)		
(First year as treatment school Indicator)* (Received treatment, when offered)					-0.012 (0.020)		
1996 district exam score, school average	0.063*** (0.021)	0.071*** (0.020)	0.063*** (0.020)	0.058 (0.032)	0.091** (0.038)	0.021 (0.026)	0.003 (0.023)
Grade indicators, school assistance controls, and time controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.23	0.23	0.24	0.33	0.36	0.28	-
Root MSE	0.273	0.272	0.272	0.223	0.219	0.150	0.073
Number of observations	56487	56487	56487	18264	18264	2327	49 (schools)
Mean of dependent variable	0.747	0.747	0.747	0.784	0.784	0.884	0.884

† The dependent variable is average individual school participation in each year of the program (Year 1 is May 1998 to March 1999, and Year 2 is May 1999 to November 1999); disturbance terms are clustered within schools. Robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Additional explanatory variables include an indicator variable for girls < 13 years and all boys, and the rate of moderate-heavy infections in geographic zone, by grade (zonal infection rates among grade 3 and 4 pupils are used for pupils in grades 4 and below and for pupils initially recorded as drop-outs as there is no parasitological data for pupils below grade 3; zonal infection rates among grade 5 and 6 pupils are used for pupils in grades 5 and 6, and similarly for grades 7 and 8). Participation is computed among all pupils enrolled at the start of the 1998 school year. Pupils present during an unannounced NGO school visit are considered participants. Pupils had approximately 3.8 attendance observations per year. Regressions 6 and 7 include pupils with parasitological information from early 1999, restricting the sample to a random subset of Group 1 and Group 2 pupils. The number of treatment school pupils from May 1998 to March 1999 is the number of Group 1 pupils, and the number of treatment school pupils after March 1999 is the number of Group 1 and Group 2 pupils.

The instrumental variables in regression 7 are the Group 1 (treatment) indicator variable, Treatment school pupils within 3 km, Treatment school pupils within 3-6 km, and the remaining explanatory variables. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

TABLE VIII

SCHOOL PARTICIPATION, SCHOOL-LEVEL DATA[a]

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)		
Panel A:					
First year post-treatment (May 1998 to March 1999)	*1st Year Treatment*	*Comparison*	*Comparison*	*Group 1 − (Groups 2 & 3)*	*Group 2 − Group 3*
Girls <13 years, and all boys	0.841	0.731	0.767	0.093***	−0.037
				(0.031)	(0.036)
Girls ≥13 years	0.864	0.803	0.811	0.057**	−0.008
				(0.029)	(0.034)
Preschool, Grade 1, Grade 2 in early 1998	0.795	0.688	0.703	0.100***	−0.018
				(0.037)	(0.043)
Grade 3, Grade 4, Grade 5 in early 1998	0.880	0.789	0.831	0.070***	−0.043
				(0.024)	(0.029)
Grade 6, Grade 7, Grade 8 in early 1998	0.934	0.858	0.892	0.059***	−0.034
				(0.021)	(0.026)
Recorded as "dropped out" in early 1998	0.064	0.050	0.030	0.022	0.020
				(0.018)	(0.017)
Females[b]	0.855	0.771	0.789	0.076***	−0.018
				(0.027)	(0.032)
Males	0.844	0.736	0.780	0.088***	−0.044
				(0.031)	(0.037)
Panel B:					
Second year post-treatment (March to November 1999)	*2nd Year Treatment*	*1st Year Treatment*	*Comparison*	*Group 1 − Group 3*	*Group 2 − Group 3*
Girls <13 years, and all boys	0.713	0.717	0.663	0.050*	0.055*
				(0.028)	(0.028)
Girls ≥14 years[c]	0.627	0.649	0.588	0.039	0.061*
				(0.035)	(0.035)
Preschool, Grade 1, Grade 2 in early 1998	0.692	0.726	0.641	0.051	0.085**
				(0.034)	(0.034)
Grade 3, Grade 4, Grade 5 in early 1998	0.750	0.774	0.725	0.025	0.049**
				(0.023)	(0.023)
Grade 6, Grade 7, Grade 8 in early 1998	0.770	0.777	0.751	0.020	0.026
				(0.027)	(0.028)
Recorded as "dropped out" in early 1998	0.176	0.129	0.056	0.120*	0.073
				(0.063)	(0.053)
Females[b]	0.716	0.746	0.648	0.067**	0.098***
				(0.027)	(0.027)
Males	0.698	0.695	0.655	0.043	0.041
				(0.028)	(0.029)

[a]The results are school averages weighted by pupil population. Standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The participation rate is computed among all pupils enrolled in the school at the start of 1998. Pupils who are present in school on the day of an unannounced NGO visit are considered participants. Pupils had 3.8 participation observations per year on average. The figures for the "Preschool–Grade 2"; "Grade 3–5"; "Grade 6–8"; and "Dropout" rows are for girls <13 years, and all boys.

[b]396 pupils in the sample are missing information on gender. For this reason, the average of the female and male participation rates does not equal the overall average.

[c]Examining girls ≥14 years old eliminates the cohort of girls in Group 1 schools (12 year olds in 1998) who were supposed to receive deworming treatment in 1998.

Gender and Human Capital

Three regularities:

A. Increasing nutrition at early ages increases the schooling of girls, not boys (e.g., Miguel and Kremer deworming).

B. Rising schooling attainment of women relative to men so that level of schooling for women exceeds that of men in most countries (e.g., United States, Bangladesh, China).

C. Higher estimated "rates of return" for women (US, almost all OECD) Psacharopoulos and Patrinos (2004): 72 of 95 countries).

What explains this?

Figure 1. Mean Years of Schooling by Gender and Urban-Rural and Year Attained Age 22, 1967-2005 (Source: 2005 Chinese Census)

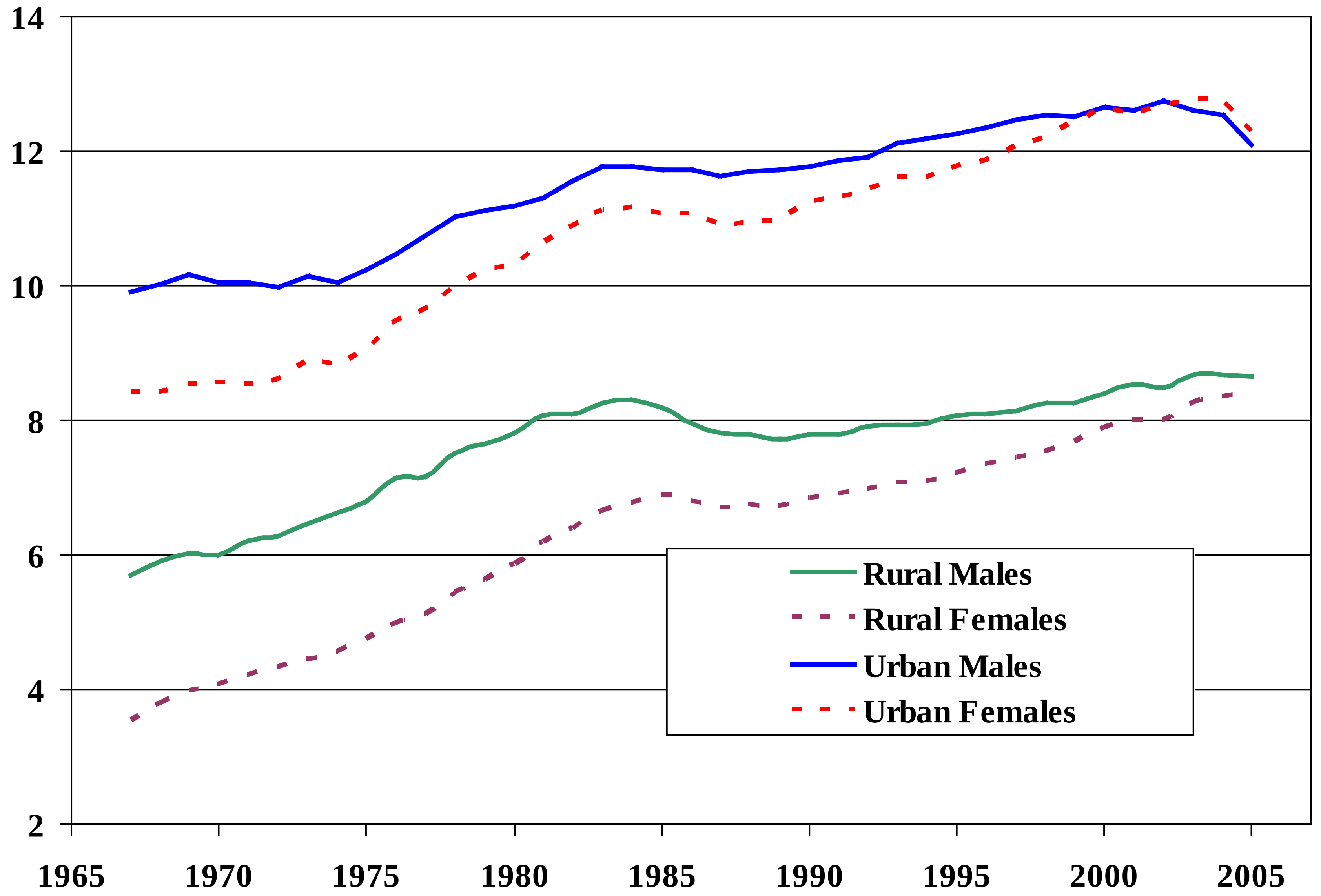
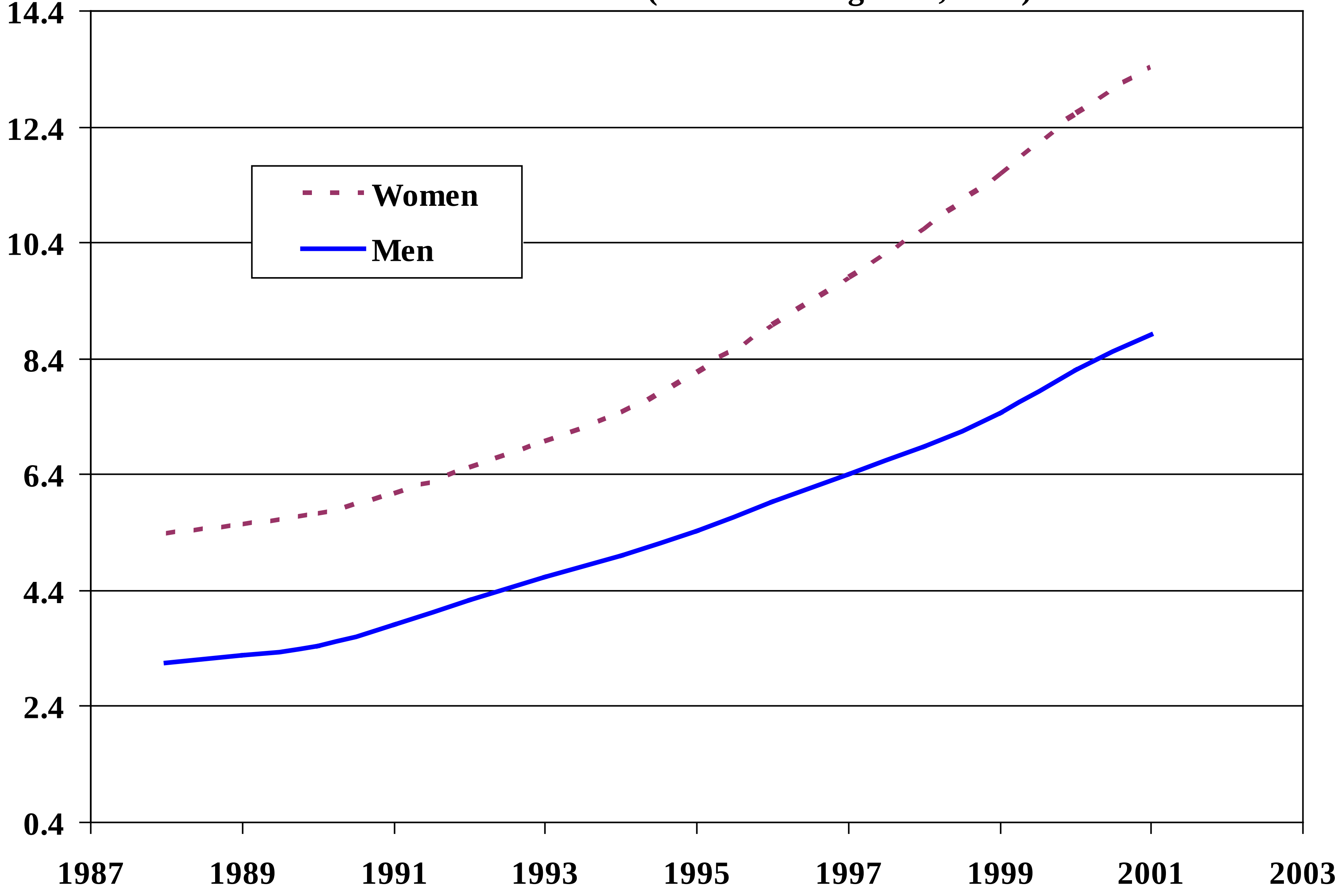
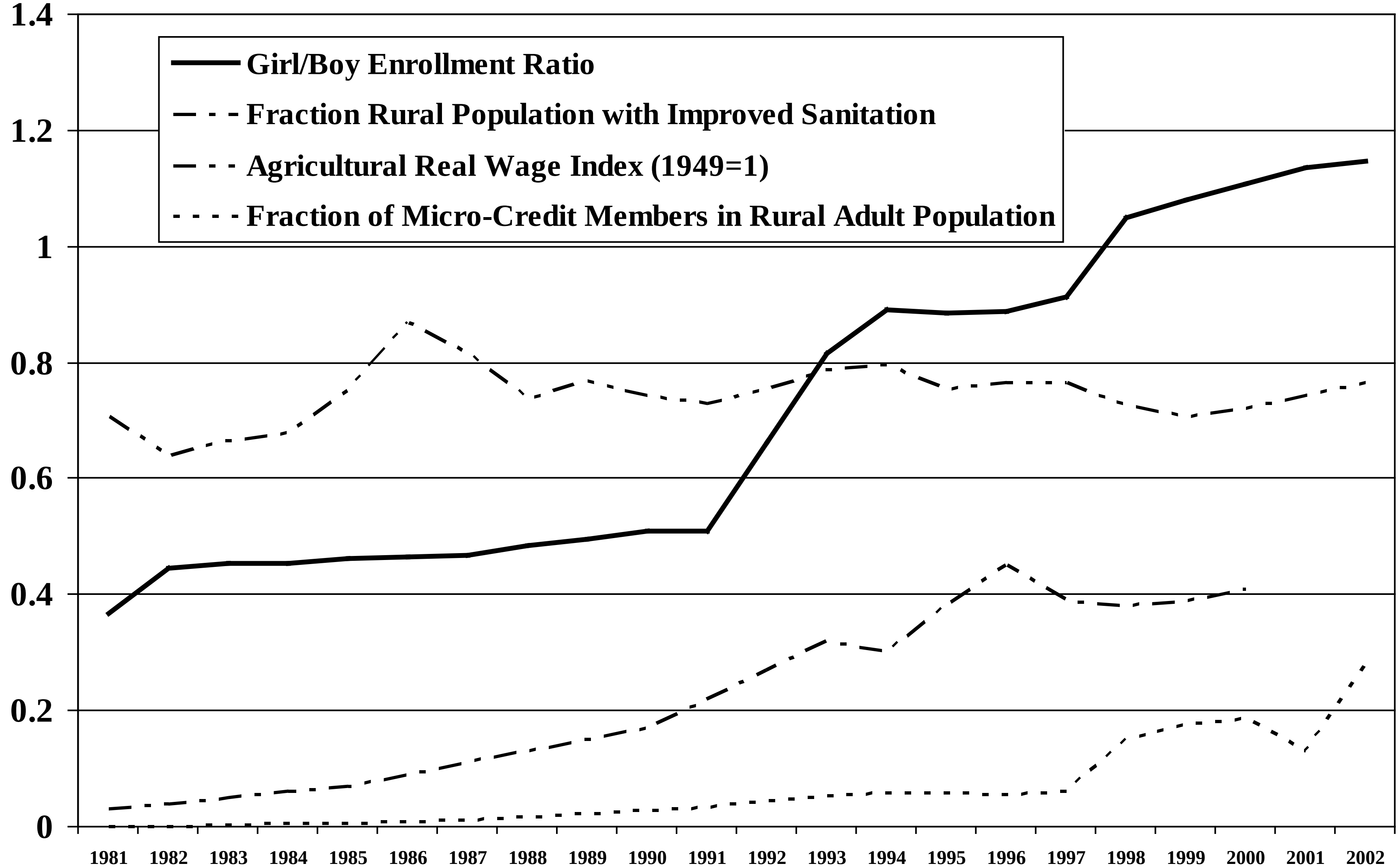
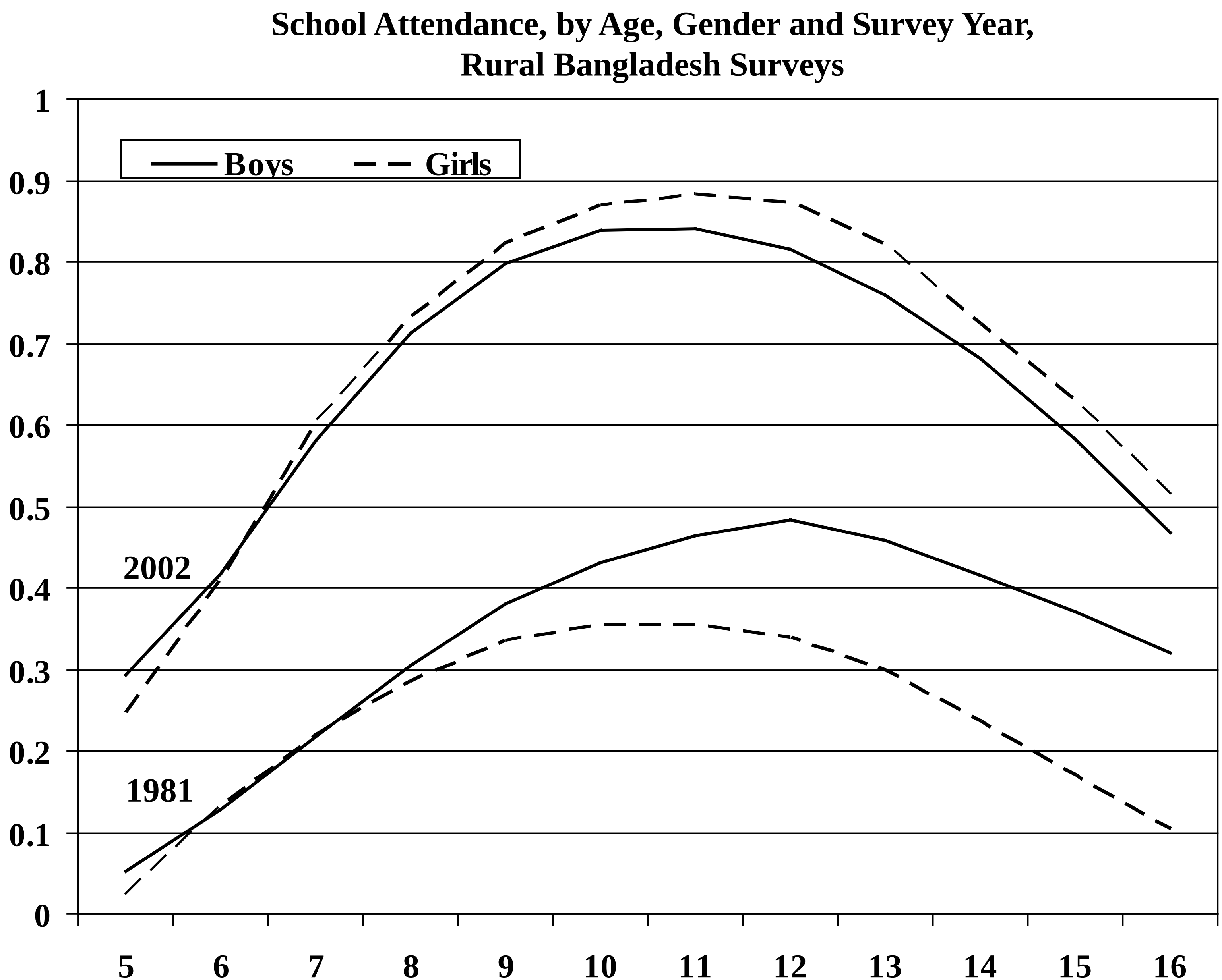


Figure 2. Estimated Rates of Return to Schooling, by Gender and Year, 1988-2001: in Five Chinese Cities (Source: Zhang *et al*., 2005)

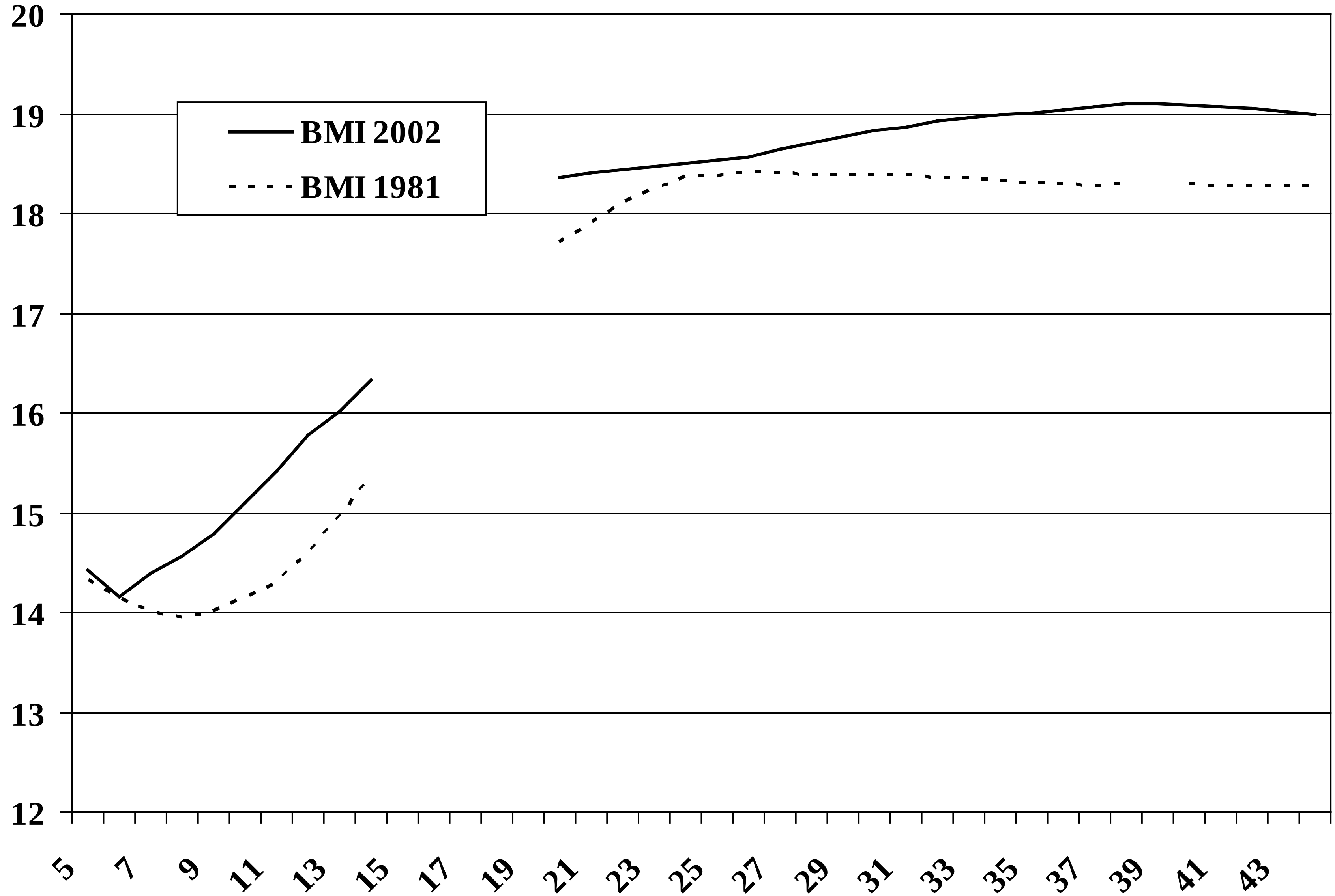


Rural Bangladesh: Ratio of Girls to Boys Enrolled in Rural Secondary Schools, Real Agricultural Wage Index, Fraction of Rural Population with Improved Sanitation, and Fraction of Adult Women Belonging to Micro-Credit Groups, 1981-2002

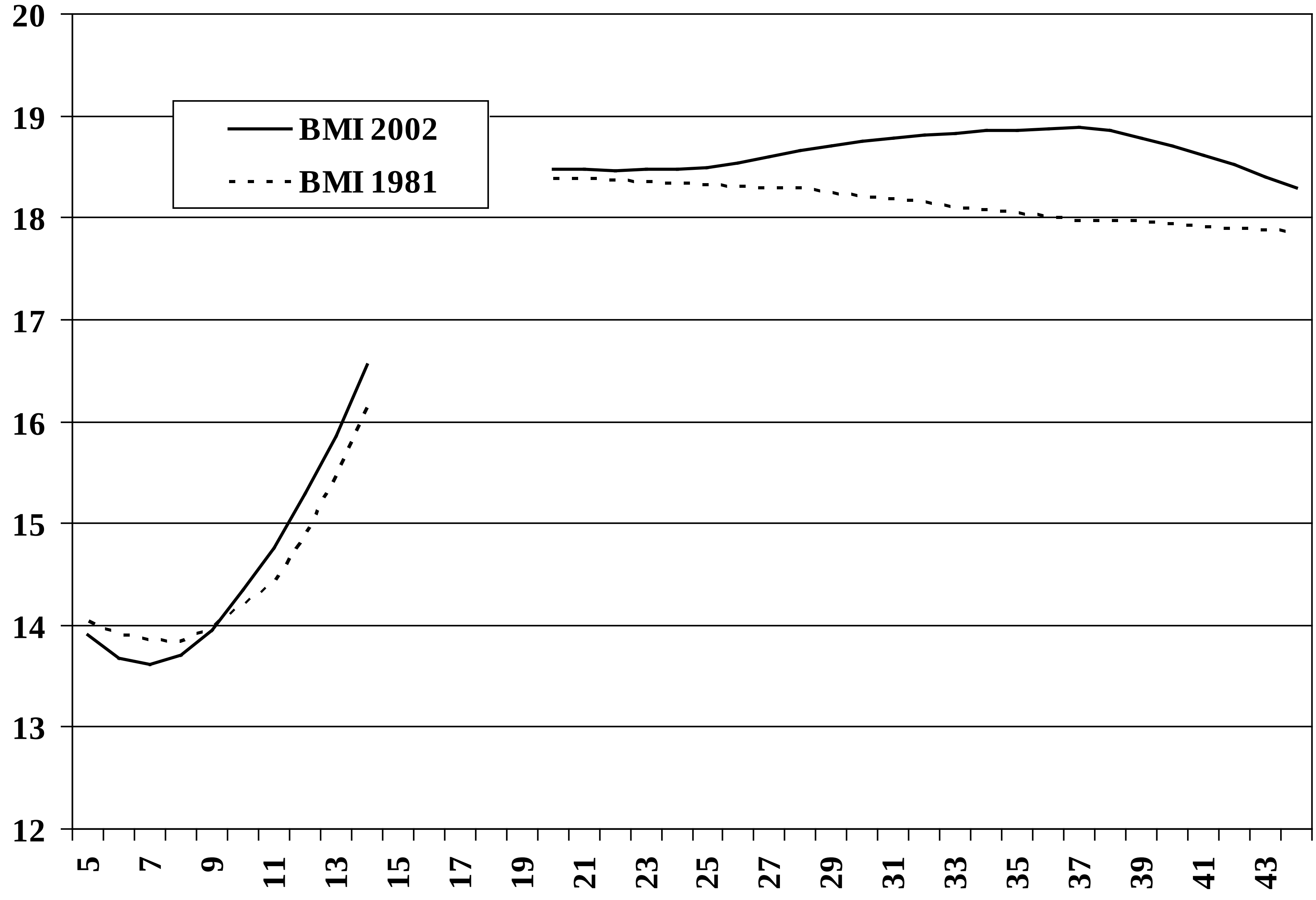



School Attendance, by Age, Gender and Survey Year,
Rural Bangladesh Surveys
Boys
Girls
2002
1981
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
6
7
8
9
10
11
12
13
14
15
16

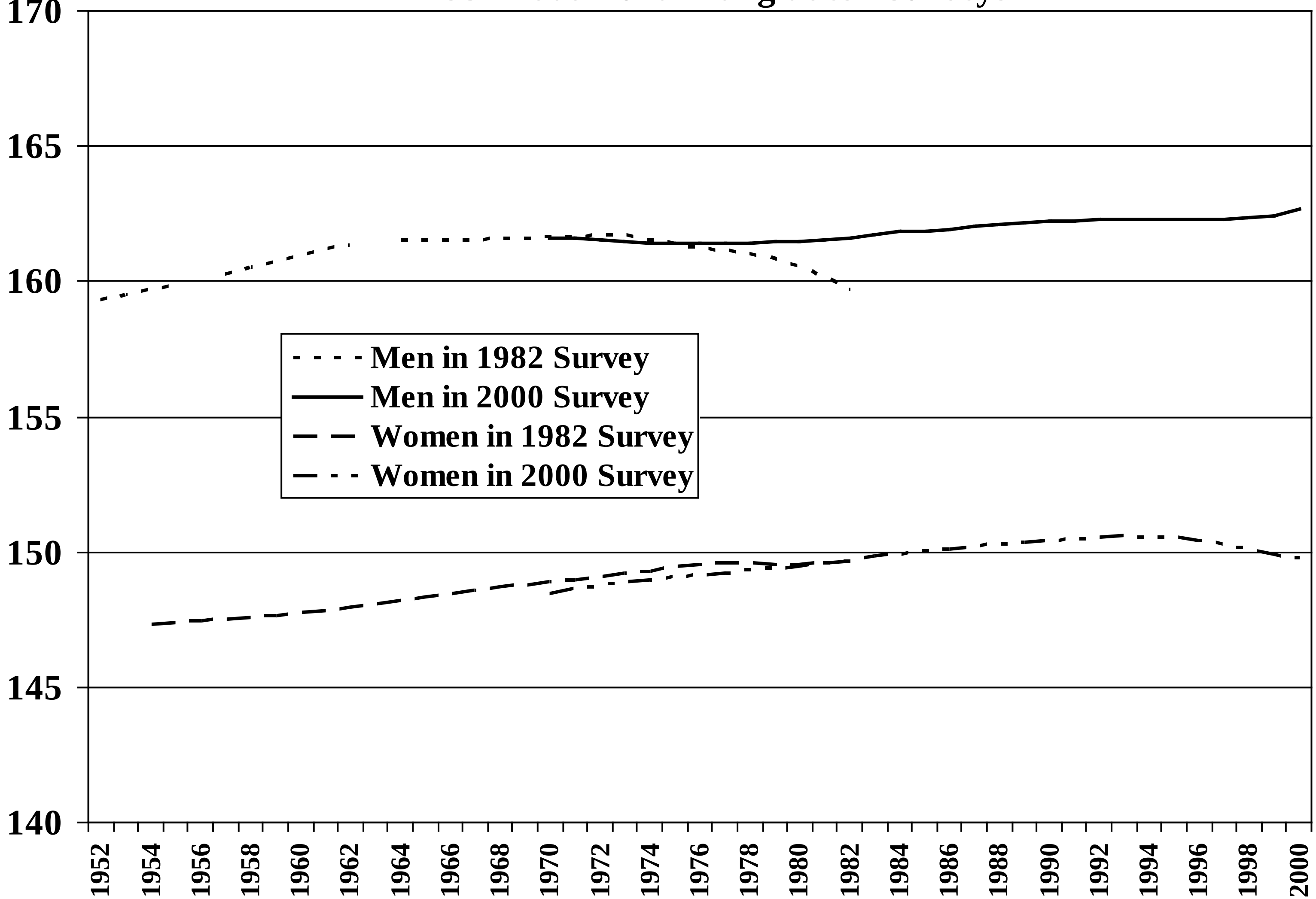
BMI, by Age and Year: Males, Rural Bangladesh Surveys



BMI, by Age and Year: Females, Rural Bangladesh Surveys



Attained Height, by Sex and Year Person Reached Age 20, 1952-2000 Rural Bangladesh Surveys



Male Occupational Distributions in Bangladesh in 2004, by Rural-Urban

Occupation/ Population	Rural	Urban
Men		
Farmer, agricultural laborer, fisherman	49.4	9.7
Unskilled laborer (rickshaw puller, brick breaking, etc.)	12.0	14.8
Factory worker or blue-collar service	3.3	7.5
Semi-skilled laborer (carpenter, mason, bus/taxi driver)	9.8	22.4
Professional (teacher, doctor, lawyer)	2.4	5.2
Business	16.4	31.3
Other	0.6	0.4
Not working	5.8	8.7

Source: Bangladesh: Demographic and Health Survey, 2004

Female Occupational Distributions in Bangladesh in 2004, by Rural-Urban

Occupation/ Population	Rural	Urban
Women		
Agricultural worker	1.2	0.4
Home-based manufacturing	3.7	3.1
Unskilled laborer (construction, brick breaking, etc.)	2.7	2.4
Poultry raising, cattle raising, trading	7.8	3.5
Domestic labor	2.0	5.8
Semi-skilled service (tailor, etc.)	3.0	6.4
Professional (teacher, doctor, lawyer)	0.5	1.7
Business	1.7	2.3
Other	0.9	1.8
Not working	76.3	72.6

Source: Bangladesh: Demographic and Health Survey, 2004

This lecture reports on work that explores the hypothesis that these phenomena reflect the *comparative* advantage of women in skill versus brawn.

Framework incorporating both occupational sorting according to comparative advantage and optimal investments by households in schooling and nutrition.

Uses unique data from rural Bangladesh and rural and urban China to test.

A. Panel data covering 25 years in rural Bangladesh with information on person-specific food consumption.

B. Child and adult twins data from urban and rural China.

The framework incorporates the role of brawn in human capital investment decisions, with specific attention to the large gender differences in brawn, to explain the gender-specific levels and trends in schooling and its returns.

Departures from previous work:

1. Incorporates brawn heterogeneity, with brawn assumed to be productive in the labor market and to affect human capital decisions.

2. Looks at nutritional investments, schooling and activity choice decisions jointly.

3. Embeds the decision model in the context of a labor market using the basic features of the Roy model (Heckman and Sedlacek, *JPE*, 1984):

Roy model features:

A. Workers are bundles of productive attributes (here, brawn, skill).

B. Workers choose activities based on their comparative advantage with respect to their attributes.

C. Rewards to attributes differ across occupations/activities: no single "rate of return" to schooling or health

4. The attributes of workers are optimally chosen.

5. The model incorporates two biological facts:

   A. Men have substantially more brawn than women.

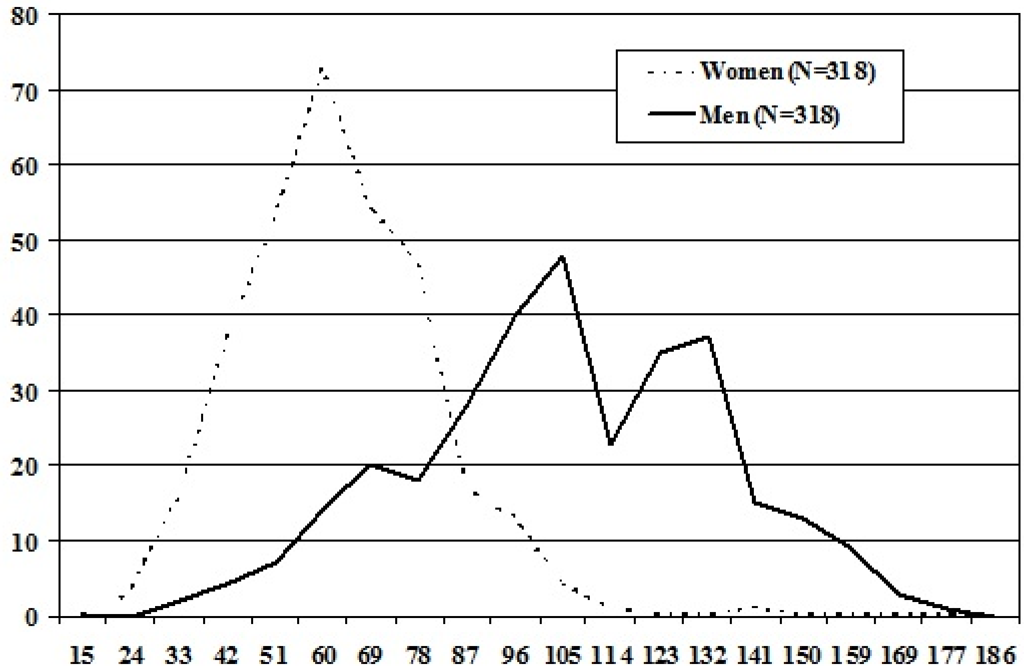
      Men thus have an absolute and a comparative advantage in brawn compared to women.

   B. Increases in nutrition increase brawn for men substantially more than for women. Medical literature: significant role of testosterone in converting nutrients to brawn.

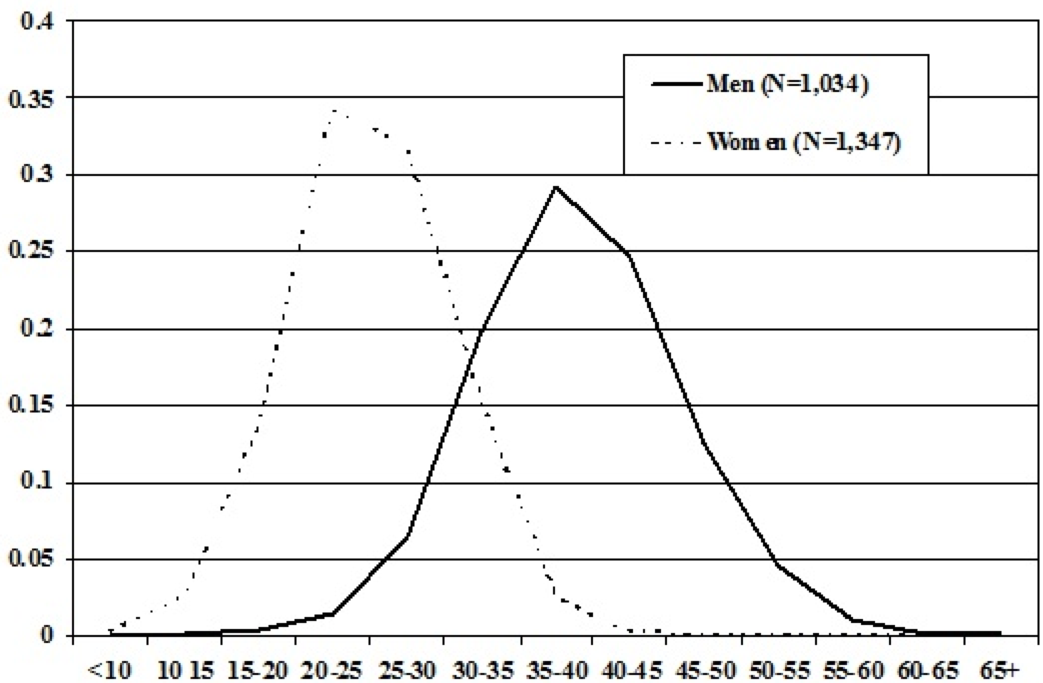
Can explore how these biological characteristics affect the response of human capital investments by gender to public health investments (Bangladesh) and changes in the skill-intensity of occupations (China).

Implications of how men and women fare under different development regimes

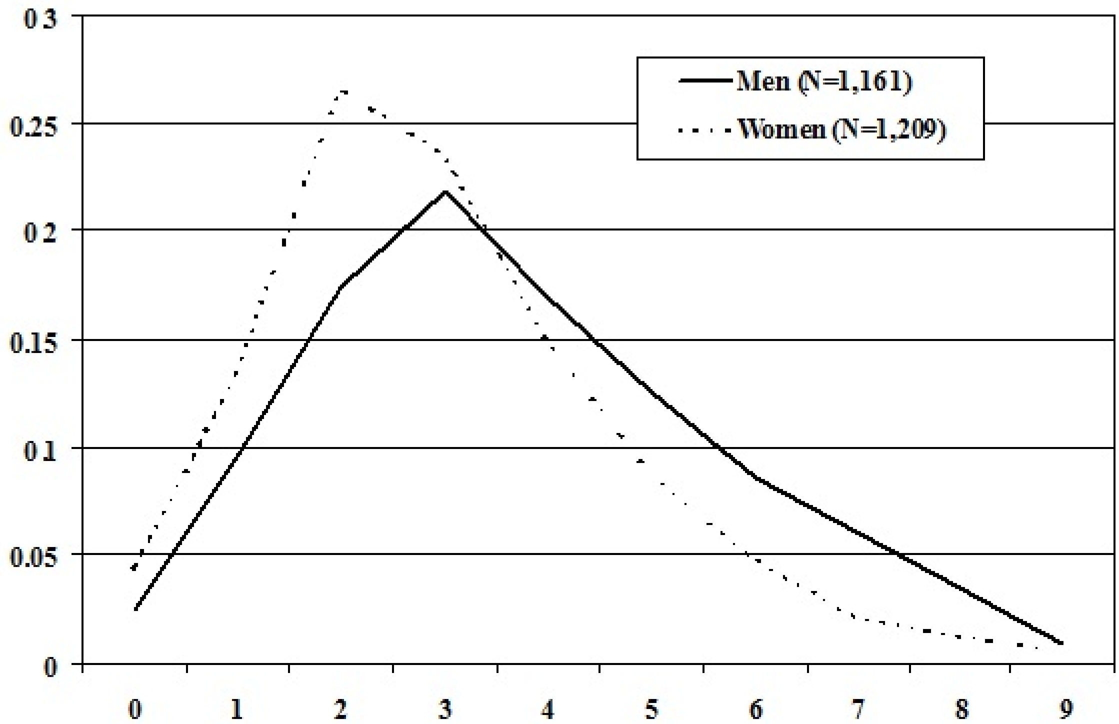
Distribution of Dynamometer Grip Strength Test Results by Gender (Mathiowetz *et. al* (1985)): Pounds of Pressure, U.S. Respondents Aged 20-94



Distribution of Dynamometer Grip Strength Test Results by Gender: Kilograms of Pressure, Respondents Aged 20-49



Distribution of Abridged Raven's Color Matrices Test Results by Gender: Number of Correct Answers, Respondents Aged 20-49



The Model

The Roy economy:

There is a continuum of tasks indexed by *i* (Ohnsorge and Treffler (*JPE*, 2007))

Each worker is a bundle of skill *H* and brawn *B*

Adult worker wage function: value of a worker's contribution to task output:

(1) $$W = \pi(i)\upsilon(i)(\kappa H)^{\alpha(i)}B^{(1-\alpha(i))}$$

where $\pi(i)$ = equilibrium price of the output of task *i*

Assumes Cobb-Douglas technology for the task function.

Order occupations/tasks by skill intensity so that $\alpha_i > 0$: thus a higher *i* means a more-skill-intensive task by definition

if $i' > i$, then $\alpha(i') > \alpha(i)$

Production technologies for brawn and skill:

(2) $$B = B(\gamma M) + b \qquad \gamma \geq 0,\ B_M > 0,\ \mathrm{B}_{\mathrm{MM}} < 0$$

Increased body mass increases brawn, at least for males

(3) $$M = M(\theta C) + m \qquad \theta > 0,\ M_1 > 0,\ M_{11} < 0$$

$\theta =$ efficiency by which calories increase body mass (lower morbidity)

Increased calorie consumption C increases body mass; there is heterogeneity in body mass endowments m

(4) $$H = H(S; \theta C) \qquad H_1 > 0,\ H_2 > 0,\ H_{12} > 0$$

Skill is augmented by schooling S and increased calorie consumption (or body mass) increases the productivity of schooling in producing skill.

(5) $$\omega = \omega(B) \qquad \omega_B > 0,\ \omega_{BB} < 0$$

The child wage is higher the higher is brawn, but is unrelated to schooling.

Choose optimal schooling, calorie consumption and activity:

The optimization program:

(6) $$\max_{C,\, S,\, i} \quad \mathrm{U}(\theta C, W)$$

(7) $$\text{s.t.} \quad F + (1 - S)\omega = pC + S\rho$$

Model results:

For men, $\gamma > 0$, the FONC are:

(8) $$\theta(\mathrm{U}_C + \gamma B_M(1 - \alpha(i))\mathrm{U}_W W/H) = \lambda[p - (1 - S)\gamma\omega_B\theta M_I B_M]$$

(9) $$\mathrm{U}_W\, \alpha(i) H_1 W/H = \lambda[\omega + \rho]$$

(10) $$\log(\kappa H/B) = -(\pi_i + v_i)/\alpha_i \pi(i)$$

(10) implies:

1. Occupation choice depends on <u>comparative</u> advantage ($\log(H/B)$)

2. If the economy is intensive in brawn, specifically $\log(H/B)<0$, then the activity price must rise as skill intensity rises - task prices are higher in skill-intensive occupations ($\pi_i>0$).

This is because, with $\log(H/B)<0$, an increase in $\alpha(i)$ lowers output, so the task price must increase to compensate.

For women, $\gamma=0$, the FONC:

(11) $$\theta \mathrm{U}_C = \lambda p$$

(12) $$U_w \alpha(i) H_1 W/H = \lambda[\omega + \rho]$$

(13) $$\log(\kappa H/B) = -(\pi_i + v_i)/\alpha_i \pi(i)$$

Implies that the returns to calorie consumption are higher for males - compare (8) to (11). Males receive more calories than females, for two reasons:

1. Calories increase the wage more for men.

2. Men work in calorie(brawn)-intensive activities, because H/B lower for men

For policy purposes, we might want to know the effects on the schooling of men and women from:

1. Reductions in morbidity - increases in θ

2. Increases in the demand for skill - the rising skill intensity of occupations α

Proposition 4: *If brawn and body mass are positively related - men, an increase in body mass may decrease schooling and the average skill-intensity of occupations.*

Proof of Proposition 4:

(19) $$\mathrm{d}S/\mathrm{d}m = (-\gamma B_1\alpha(i)H_1[(1 - \alpha(i))W/HB + \mathrm{U}_{WW}W/H] + \lambda\gamma\omega_B B_1)\Phi_{22} - (\gamma B_1/B)\Phi_{32}$$

(20) $$\mathrm{d}i/\mathrm{d}m = (-\gamma B_1\alpha(i)H_1[(1 - \alpha(i))W/HB + \mathrm{U}_{WW}W/H] + \lambda\gamma\omega_B B_1)\Phi_{32} + (\gamma B_1/B)\Phi_{33}$$

Lemma 1: *If brawn and body mass are positively related only for males, then increases in body mass for everyone will decrease schooling for males relative to females and increase the gender division of labor* (*difference in average* $\alpha(i)$ *between men and women*).

Lemma 2: *If men have more brawn than women, both the amounts of schooling and the "returns" to schooling will be higher for women than for men.*

The Data: **Bangladesh**

Three rounds of data describing households in rural Bangladesh:

1. *Nutrition Survey of Rural Bangladesh* 1981-82 N=4,107

    Probability sample of 50 households in each of 15 villages meant to be representative of the rural population of Bangladesh

2. *Nutrition Survey of Rural Bangladesh* 2002 N= 9,838

    Includes all individuals surveyed in 1981 and their households in 14 of the 15 fifteen villages + a new random sample of households in the 14 villages

    Attrition of surviving individuals = 3% (individuals included no matter where they reside in Bangladesh)

3. *Nutrition Survey of Rural Bangladesh* 2007-2008 N=12,244

    Includes all individuals surveyed in 2002 and their households

Important features of the Bangladesh data

A. <u>Individual-specific</u> food intake measured over a 24-hour period by observation and measurement for all individuals in each round (first round = random ½ of households)

B. Individual-specific activity schedules

C. Individual anthropometric information: height, weight, arm circumference

D. Households in two villages interviewed multiple times in the same year:

   Validation subsample: 4 repetitions in 1981-82; 2 repetitions in 2002 and 2007-8 (provide exchangeable replicates)

E. Individual-specific assessments of grip strength, pinch strength and aptitude (Raven's matrices, digit span) in 2007-8

Features A-C enable the estimation of body mass production functions and thus body-mass endowmenmts

Feature D enables correction for measurement error in endowment measures

Estimation Strategy using the Bangladesh Data

The main objectives are:

1. To estimate the effects of body mass endowments for males (brawn) and females on schooling choice and occupation selection and on wages for men

2. To estimate the wage function incorporating (a) the body-mass endowment (brawn) and (b) heterogeneity in the returns to skill and brawn across occupations *i* consistent with the Roy model

Step 1: Estimate the body-mass production function (3), using the same methodology as in Pitt *et al*. (*AER*, 1990): weight/height is outcome

Challenge: inputs (C) are endogenous (also include activity levels)

IV estimates: instruments are village-level price interacted with individual age, household land holdings and head's characteristics - age and schooling.

Note: repeated cross-sections (20+ years apart) enables testing whether *structural* (biological) estimates, immune to environmental change

Step 2: Estimate r-f endowment effects using residual m from production function

(22) $$y_j = \mathbf{Z}_j\zeta + bm_j + \varepsilon_j \qquad\qquad y_j = S_j, W_j, i_j$$

Challenge: the residual m_j for individual j contains the true body mass m_j*, net of the influence on body mass of current consumption and activities, + measurement error η_j

$$m_j = m_j^* + \eta_j$$

Both the vector of coefficients ζ and b would be biased

Use the validation sample of within-round replicates: assume classical measurement error properties for η_j and repeated measures have the same mean and independent measurement errors.

For the validation subsample (households in two villages) we have

(23) $$m_{jr} = m_j^* + \eta_{jr} \qquad\qquad \text{where r=1,2,3,4 or 1,2}$$

Jointly estimate the outcome equation (22) and the measurement equation (23) using maximum-likelihood (GLLAMM)

Table 1
2SLS Estimates of the (Cobb-Douglas) Body-Mass Production Function, by Survey Population

Dependent variable: log weight/height		
Input/Survey population	1981	2002
Log individual total calorie consumption[a]	.136 (3.37)	.241 (3.76)
Very active occupation[a]	-.0119 (0.23)	-.0445 (3.20)
Exceptionally active occupation[a]	-.0817 (1.26)	-.125 (5.65)
Pregnant[a]	.326 (1.34)	.0273 (1.33)
Lactating[a]	.513 (4.65)	.0339 (1.39)
Log age	.0987 (1.90)	.00804 (9.02)
Log age squared	.0174 (2.37)	-.000092 (8.86)
Male	-.0578 (1.81)	-.00947 (2.89)
Male*log age	.0687 (4.04)	.00116 (3.72)
Water drawn from tube well	-.0406 (2.10)	.00551 (0.88)
Water drawn from well	-.0693 (3.15)	.00118 (0.18)
Water drawn from pond	-.0649 (2.55)	.0216 (2.36)
N	1,737	5,750
H_0: calories, age, age squared, male, exceptionally active, male*age = across populations; $\chi^2(6)$ (p)	9.06 (.17)	

Sources: NSRB 1981 and 2002. [a]Endogenous variable: instruments include household head's age and schooling level, land holdings, and price of all foods consumed interacted with individual age and sex variables, land and head's schooling and age. Asymptotic *t*-ratios in parentheses.

Table 3

The Body-mass Endowment and Grip Strength (2007), by Gender: Respondents Aged 20-49 in 2002

Dependent variable: Kilograms of pressure				
Group	Men		Women	
Estimation procedure	GLS	GLLAM	GLS	GLLAM
Endowment	7.41	8.17	1.58	1.93
	(5.89)	(6.03)	(1.69)	(1.67)
Household owned land	.00223	.00235	.000500	.000507
	(1.74)	(2.11)	(0.46)	(0.53)
No land owned	-.603	-.546	-.433	-.433
	(1.18)	(1.08)	(1.12)	(1.15)
Age	-.389	-.330	.177	.174
	(1.56)	(1.32)	(0.92)	(0.93)
Age squared	.00347	.00266	-.00520	-.00517
	(0.97)	(0.73)	(1.87)	(1.90)
ρ	-	.907	-	.796
N	946	946	1,087	1,087

Source: NSRB 2002-2007 panel. Absolute values of asymptotic *t*-ratios corrected for clustering within households in parentheses. All specifications include village fixed effects.

Table 4
The Body-mass Endowment in 1982 and Completed Schooling in 2002, by Gender:
Respondents Aged 0-15 in 1982

Dependent variable: Completed education (years) in 2002				
Gender	Male		Female	
Estimation procedure	GLS	GLLAM	GLS	GLLAM
Endowment (1982)	-1.82	-2.22	1.00	.923
	(2.05)	(1.94)	(0.86)	(0.67)
Age	-.164	-.161	-.110	-.123
	(0.76)	(0.77)	(0.49)	(0.57)
Age squared	.00501	.00483	-.00908	-.00787
	(0.36)	(0.36)	(0.60)	(0.54)
Household owned land (1982)	.00711	.00711	.00467	.00468
	(5.49)	(5.66)	(3.05)	(3.17)
No land owned (1982)	-1.02	-1.03	-1.05	-1.05
	(1.48)	(1.54)	(1.55)	(1.68)
Average household endowment	.261	.270	.116	.158
(1982)	(0.12)	(0.17)	(0.06)	(0.08)
ρ	-	.838	-	.821
N	311	311	273	273

Source: NSRB 1982-2002 panel. All specifications include village fixed effects; clustered t-ratios.

Table 5
The Body-mass Endowment in 1982 and Log Wages in 2002: Males Aged 0-17 in 1982

Dependent variable: Log daily wage in 2002 (*tk.*)						
Estimation procedure	GLS	GLLAM	GLS	GLS	GLLAM	GLLAM-IV
Schooling (2002)	-	-	.0261 (2.27)	.0277 (2.41)	.0311 (3.01)	.0432 (2.18)
Endowment (1982)	.252 (1.21)	.327 (2.16)	-	.303 (1.46)	.370 (2.49)	.387 (2.39)
Age	.100 (2.30)	.106 (2.73)	.0750 (1.90)	.0994 (2.33)	.103 (2.66)	.100 (2.57)
Age squared	-.00517 (2.08)	-.00536 (2.34)	-.00356 (1.61)	-.00510 (2.08)	-.00527 (2.30)	-.00518 (2.26)
Household owned land (1982)	.000404 (2.03)	.000412 (2.51)	-	-	-	-
No land owned (1982)	.0528 (0.44)	.0607 (0.68)	-	-	-	-
Average household endowment of others (1982)	-.115 (0.27)	-.147 (0.45)	-	-	-	-
ρ	-	.838	-	-	.838	.838
N	225	225	225	225	225	225

Table 8
Estimated Marginal Effects of the Body-Mass Endowment on the Probability of Attending School, by Gender and Estimation Method:
Children Ages 10-15 in 2002

Estimation method	Multinomial Logit[a]		ML Logit-GLLAM[b]	
Group	Boys	Girls	Boys	Girls
Endowment	-.248 (2.37)	.0600 (0.68)	-.436 (3.60)	.0983 (1.04)
Household land owned	.00670 (3.23)	.00135 (1.14)	.00674 (3.42)	.00135 (1.11)
No land owned	-.0417 (1.08)	-.0463 (1.64)	-.0420 (1.24)	-.0465 (1.70)
Household average endowment of other family members	-.0955 (0.62)	.0667 (0.87)	-.0927 (1.10)	.0784 (1.09)
Age	-.0332 (3.43)	-.0243 (2.54)	-.0325 (3.23)	-.0236 (2.53)
N	410	353	410	353

Source: NSRB 2002. Source: NSRB 2002. [a]Asymptotic *t*-ratios corrected for clustering at the household level in columns. [b]Bootstrapped *t*-ratios in parentheses in columns.

Creating an Occupational Index based on Activity-Specific Energy Requirements

From Annex 5 "Energy Costs of Activities," from *Human Energy Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation*, Rome, 17–24 October 2001.

Energy requirement = the amount of food energy needed to balance energy expenditure in order to maintain body size in a given activity.

Data are ratio of average energy requirements per unit of time divided by the basal metabolic rate (energy expenditure at rest) by activity:

Examples: PAR's (physical activity rates):

Pulling a rickshaw with two passengers	7.2
Weeding	4.0
Sawing hardwood	6.6
Filing, reading, writing	1.3
Bed making	3.4

We used these PAR's, and adult male BMR's to calculate energy expenditures by activity per kilogram = *kilojoules/kilogram per day*

Table 9
The Body-mass Endowment and Occupation Choice, by Gender: Respondents Aged 20-49 in 2002

Dependent variable: Occupational Energy Expenditure

Group	Men		Women	
Estimation procedure	GLS	GLLAM	GLS	GLLAM
Endowment	81.7	93.6	-5.78	-6.78
	(14.9)	(11.6)	(3.62)	(2.58)
Age	.969	.957	1.05	1.06
	(0.89)	(0.90)	(3.23)	(2.86)
Age squared	-.00153	-.000825	-.0132	-.0133
	(0.10)	(0.05)	(2.80)	(2.63)
Household owned land	-.0195	-.0212	-.000269	-.000262
	(3.13)	(3.42)	(0.15)	(0.18)
No land owned	6.62	6.22	2.37	2.38
	(2.87)	(2.67)	(3.64)	(3.11)
Average household endowment	-27.0	-29.0	-2.02	-2.02
of others (1982)	(4.14)	(4.53)	(1.10)	(.91)
ρ	-	.901	-	.828
N	1,236	1,236	1,338	1,338

Source: NSRB 2002. Absolute values of asymptotic *t*-ratios corrected for clustering within households in parentheses. All specifications include village fixed effects.

Step 3: Estimate a wage function in terms of schooling and body mass consistent with the Roy model structure

(1) $$W(i)_j = \pi(i) H_j^{\alpha(i)} B_j^{(1 - \alpha(i))} \xi_j$$

Challenges: Schooling is endogenous, the body mass residual measure contains measurement error, and the coefficients are functions of endogenous occupation. Indeed, the Roy model implies a different wage function ($\alpha(i)$) for each job, with job choice endogenous.

To solve the latter problem, assume that skill-intensity $\alpha(i)$ is inversely related to the energy expenditure $\varepsilon(i)$ of an occupation and $\pi(i)$ is thus aslo a function of $\varepsilon(i)$.

In particular:

$$\alpha(i) = \alpha_0 + \alpha_1 \varepsilon(i) \qquad \alpha_0 > 0,\ \alpha_1 < 0$$

$$\pi(i) = \varepsilon(i)^{\delta}$$

Let the skill and brawn production functions be given by:

$$H_j = e^{(\beta S(j)\ +\ \varsigma \mathrm{age(j)})} \qquad \text{and} \qquad B_j = e^{\gamma m(j)} \qquad \beta>0,\ \gamma>0$$

Then the estimable wage function in terms of the structural parameters is:

$$\log W(i)_j = \delta \log\varepsilon(i) + \alpha_0 \beta S_j + \alpha_1 \varepsilon(i)\beta S_j + (1 - \alpha_0)\gamma m_j - \alpha_1 \varepsilon(i)\gamma m_j$$

$$\alpha_0 \varsigma age_j + \alpha_1 \varepsilon(i)\varsigma \mathrm{age}_j + \log\xi_j$$

And the estimating equation is:

$$\log W(i)_j = \delta \log\varepsilon(i) + v_1 S_j + v_2 \varepsilon(i) S_j + v_3 m_j - v_4 \varepsilon(i) m_j$$

$$v_5 age_j + v_6 \varepsilon(i)\mathrm{age}_j + \log\xi_j$$

Model sign restrictions: $v_1>0,\ v_2<0,\ v_4>0$

and the sign of δ indicates whether the economy is brawn-based:

$$\delta<0 \text{ if } (\log(H/B)<0$$

Use GLAMM-IV to estimate: measurement error in m_j
endogenous S_j, ε_j
bootstrap standard errors

For women:

1. Only 79 women work for wages, almost all as domestic workers.

2. Selection-corrected wage function

3. Theory says parameters of structural (Roy) wage function should be the same across genders, but may not be if support of the activity distribution is small

Table 9
Occupation-Specific Wage Function Estimates, by Gender: Adults Aged 20-49 in 2001-2

Gender	Male			Female		
Estimation procedure	GLS[a]	GLLAM[a]	GLLAM-IV[b]	GLS[a]	GLLAM[a]	GLLAM-IV[b]
Schooling	.0409 (11.6)	.0417 (10.1)	.334 (2.75)	.0487 (2.38)	.0467 (2.41)	1.14 (2.05)
Schooling x occupation energy expenditure	-	-	-.00256 (2.87)	-	-	-.007 (2.03)
Endowment	-	.0765 (0.84)	-1.46 (2.34)	-	.0895 (0.22)	-4.53 (1.28)
Endowment x occupation energy expenditure	-	-	.0115 (2.85)	-	-	.0254 (1.52)
Age x occupation energy expenditure	-	-	.000401 (1.16)	-	-	.00204 (1.09)
δ	-	-	-3.36 (1.67)	-	-	-4.37 (0.44)
λ	-	-	-	11.7 (1.46)	12.7 (1.49)	10.3 (0.78)

ρ	-	.889	.867	-	.904	.900
Sargan overid. test χ^2 [p]	-	-	11.09 [.436]	-	-	.0033 [.991]
N	1,094	1,094	1,094	79	79	79

Source: NSRB 2001-2. All specifications include age age and age squared. [a]Absolute values of asymptotic *t*-ratios corrected for clustering within households in parentheses in column. [b]Bootstrapped *t*-ratios in parentheses in column.

Wage function estimates indicate that:

1. The returns to schooling are lower in energy-intensive occupations; the returns to schooling in less energy-intensive occupations.

In particular, the effect of increased body-mass on the wage is positive for activities engaged in by 64% of male workers.

Schooling has a positive return for low-energy-intensive activities; these include clerks and tailors, in which women are over-represented.

2. The estimate of δ is negative - high energy-intensive activities are priced lower than low enery-intensive activities, consistent with male workers overall having a comparative advantage in brawn.

Empirical analysis using Bangladesh data has two shortcomings:

1. Use of residual measure of health endowments relies on obtaining correct estimates of health production function; functional form

   Data on twins enables us to estimate endowment effects without functional form assumptions - non-parametric estimates

2. Estimates obtained in static environment: no ability to see how economic development affects relationships

   A. Change in occupational structure; e.g., higher skill intensity α

   B. Increases in family income

China Survey Data

Samples for analysis from three surveys:

1. The *Chinese Twins Survey* (CTS), carried out by the Urban Survey Unit (USU) of the National Bureau of Statistics (NBS) in June and July 2002 in five cities of China: Chengdu, Chongqing, Harbin, Hefei, and Wuhan.

   The local Statistical Bureaus identified same-sex twins aged between 18 and 65. N = 1,495 matched pairs of twins (2,990 respondents).

   Sample for analysis: 611 male and 326 female twins pairs aged 18-29 (schooling in post reform era) [12 birth cohorts in each of 5 cities]

   Information provided on birthweight, current height and weight, educational attainment, occupation, and earnings.

2. The *CTS non-twins survey*: a probability sample of 1,665 non-twin individuals aged 25-60 in the same five cities based on the sample frame of the regular urban household surveys. Same questionnaire.

3. The *Chinese Child Twins Survey* (CCTS), carried out by the Urban Survey Unit (USU) of the National Bureau of Statistics (NBS) in late 2002 and early 2003 in the Kunming district of China.

   Households with child twins aged between 7 and 18 in both rural and urban areas. N = 1,694 twin pairs.

   Sample for analysis: 194 male and 222 female rural twin pairs and 205 male and 211 female urban twin pairs aged *12-15*

The CTS enables the identification of endowment effects under different occupational regimes

Ordinarily it is not possible to reconstruct a time-series of the aggregate occupational structure based on the current occupations of different age cohorts in a single cross-section.

The *non-twins* CTS provides the occupation of each respondent when he or she was first married and the year of that marriage.

Almost all respondents married, and almost all married within the age range 20-29

We use the at-marriage information to construct a time series of occupations by year, based on year of marriage, for young workers.

Occupation Categories in China Census and UHS

1. Professional and technical (doctor, professor, lawyer, architect, engineer, etc.)

2. Administrator/executive/ cadre (working proprietor, government official, section chief, department or bureau director, administrative cadre, etc.)

3. Office staff member

4. Commerce staff

5. Service worker

6. *Agricultural worker*

7. *Manufacturing worker , transport worker and other skilled and non-skilled workers (foreman, driver, sailor, ordinary labor)*

8. Others

Figure 3. Proportion of Employment in in Non-Brawn Occupations, by Year, 1968-2002: in Five Chinese Cities (Source: 2002 Adult Nontwin Survey)

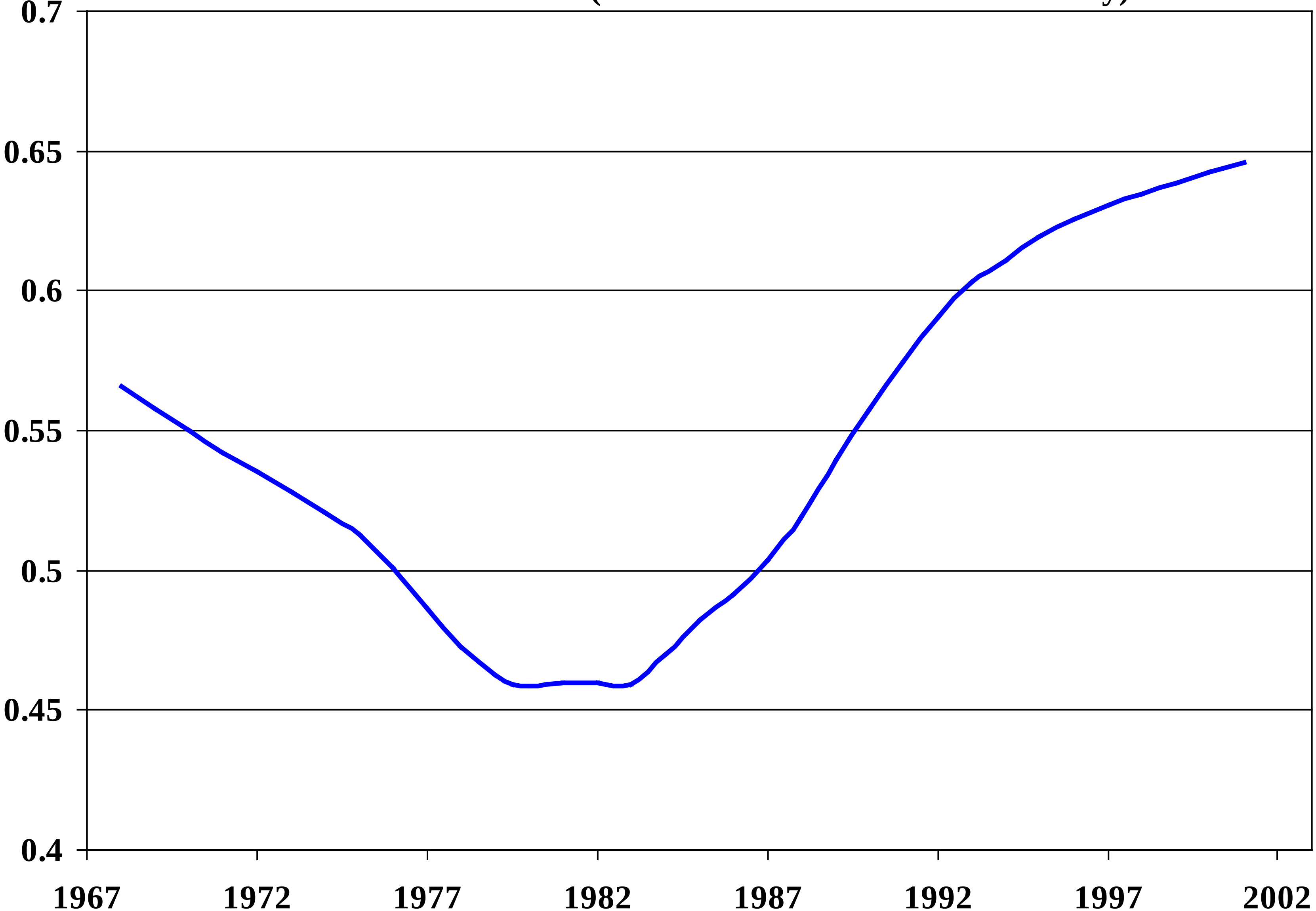
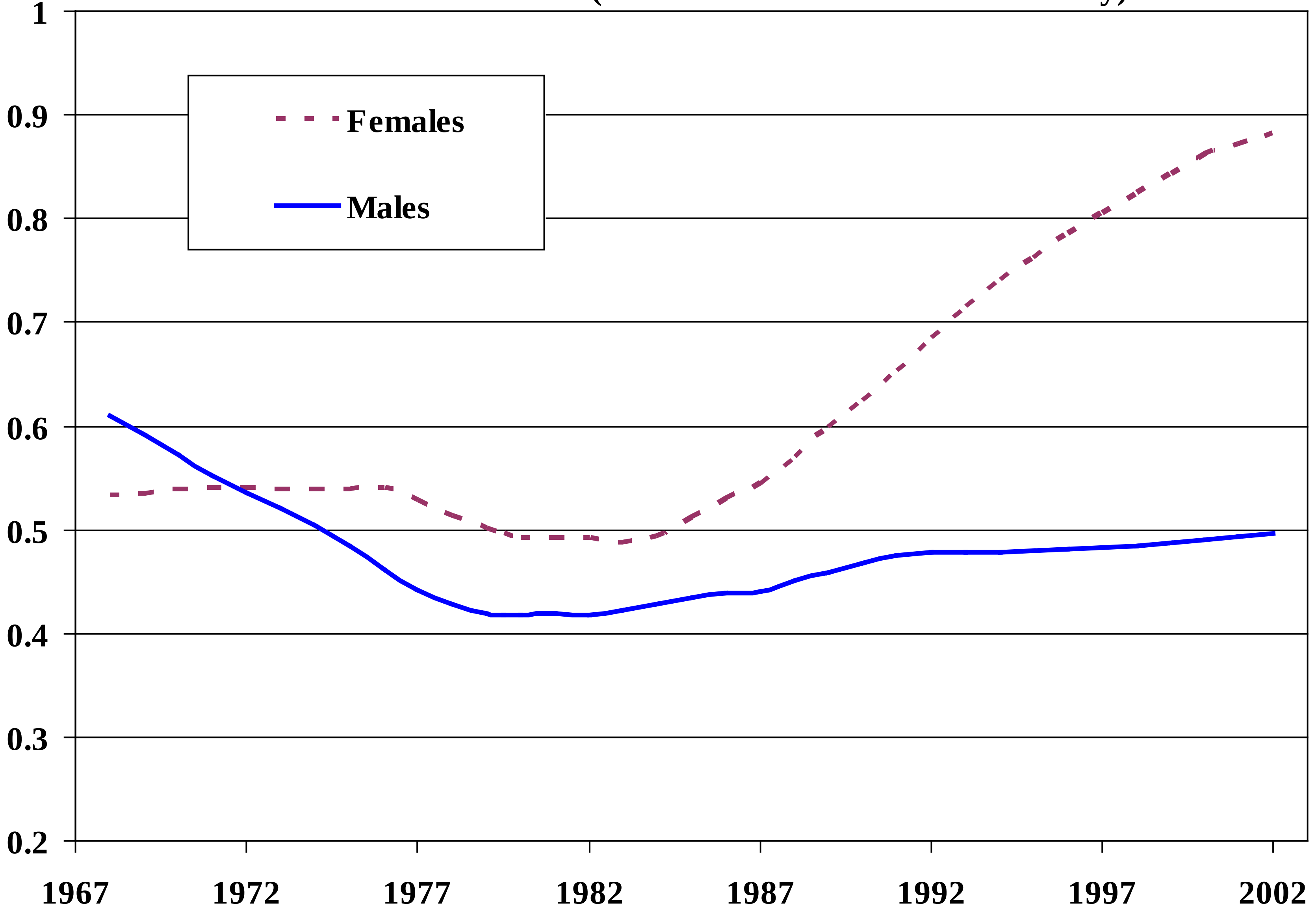


Figure 4. Proportion of Employment in in Non-Brawn Occupations, by Gender and Year, 1968-2002: in Five Chinese Cities (Source: 2002 Adult Nontwin Survey)



With the twins data sets, we estimate the differenced (within-twin pair) versions of the reduced form equations.

We also estimate the endowment reduced-form equations non-parametrically, allowing the effects to differ by family income and α:

(17) $$\Delta S_{ijk} = \beta_{kF}\Delta m_{ijk} + \Delta\varepsilon_{ijk}, \quad k\text{=m,f}$$ Child twins (CCTS)

(18) $$\Delta S_{ijk} = \beta_{k\alpha}\Delta m_{ijk} + \Delta\eta_{ijk}, \quad k\text{=m,f}$$ Adult twins (CTS)

Figure 2
Distribution of the Absolute Value of Birthweight Differentials (oz.) Among identical Twins

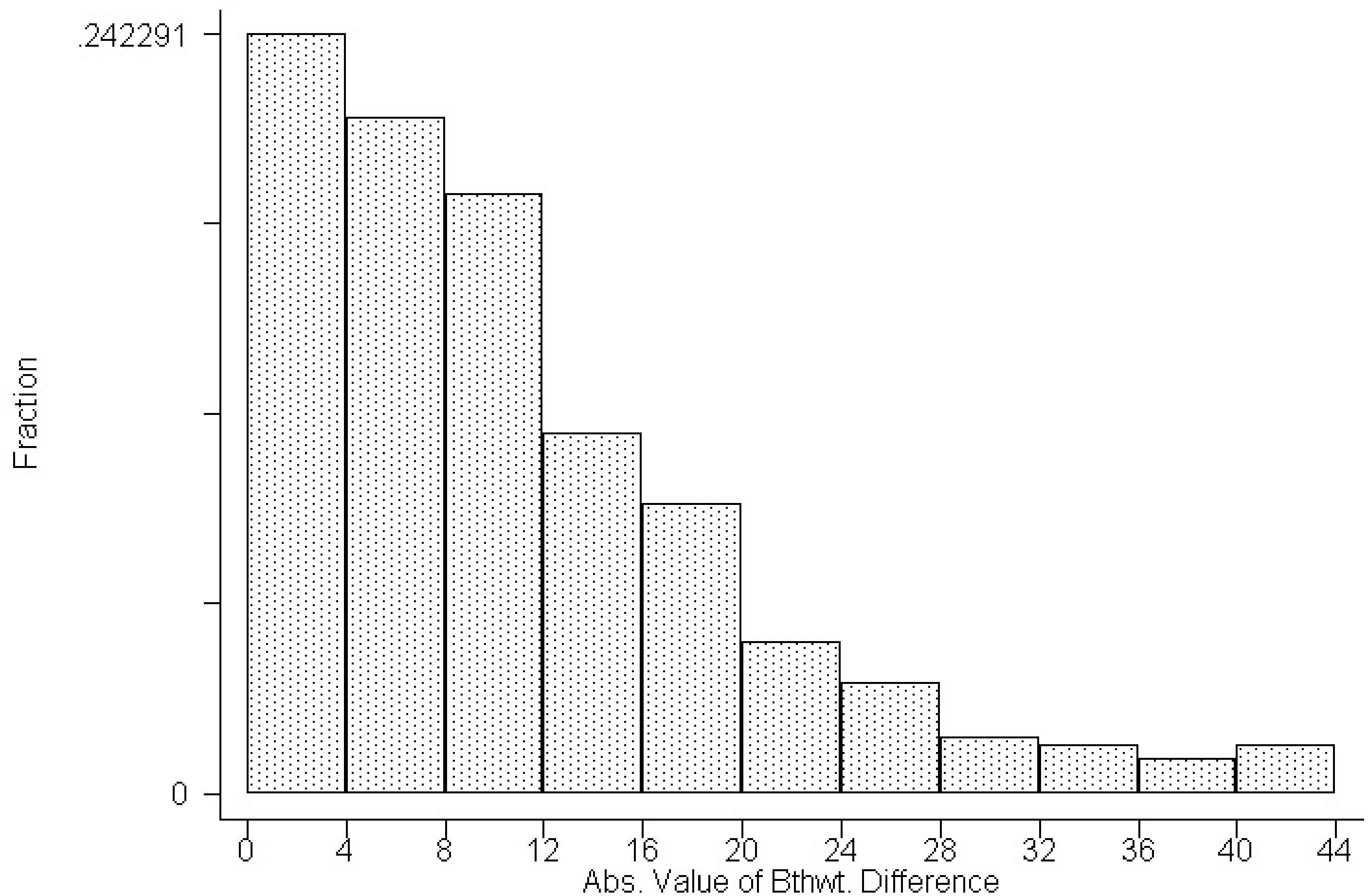


Table 4
Within-twin Estimates of Birthweight Effects on Schooling Attainment, Wages, and Weight-for-Height in Urban China, by Gender: Twins Aged 18-29
(Source: Chinese Adult Twins Survey, 2002)

	Schooling Attainment (Years)		Log Monthly Wage		Weight for Height	
	(1)	(2)	(3)	(4)	(5)	(6)
Birthweight	.276 (1.08)	-	.111 (1.62)	-	.0093 (1.95)	-
Birthweight - female	-	.867 (2.17)	-	.0543 (0.52)	-	.00134 (0.18)
Birthweight - male	-	-.0275 (0.09)	-	.195 (2.46)	-	.0133 (2.37)
t-statistic: difference male - female [p]	-	-1.92 [.052]	-	2.08 [.038]	-	1.36 [.175]
Number of twins	936	937	744	744	936	936

Asymptotic t-ratios in parentheses.

Table 5
Within-twin Estimates of Birthweight on Educational and Health Outcomes,
by Gender and Rural-Urban Location: Twins Aged 12-15
(Source: Chinese Child Twins Survey, 2002)

	Mean Language and Math Grade (Percent)			Number of Student Honors per Year			Weight for Height		
	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban
Birthweight - female	5.55	6.44	4.63	.0526	.0988	.00679	.0097	.0075	.0113
	(2.88)	(2.23)	(1.79)	(1.57)	(2.31)	(0.13)	(2.66)	(1.31)	(2.44)
Birthweight - male	3.69	4.88	2.28	.0187	.0705	-.0376	.0108	.0097	.0107
	(2.00)	(1.79)	(0.91)	(0.58)	(1.75)	(0.74)	(3.08)	(1.80)	(2.39)
t-statistic: difference male - female [p]	-3.77	-1.92	-3.22	-3.94	-2.80	-3.03	1.12	1.66	0.40
	[.000]	[.057]	[.002]	[.000]	[.006]	[.003]	[.264]	[.099]	[.686]
F(2, 414): urban = rural [p]	-	0.570 [.566]		-	1.88 [.154]		-	1.09 [.338]	
Number of twins	830	414	416	832	416	416	832	416	416

Asymptotic t-ratios in parentheses.

Table 6
Within-twin Estimates of Birthweight on Parental Education and Health Expectations, by Gender and Rural-Urban Location: Twins Aged 12-15
(Source: Chinese Child Twins Survey, 2002)

	Expected Years of Schooling Completed			Expect Attend College			Expect Good health		
	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban
Birthweight - female	.400	.504	.318	.103	.112	.107	.116	.140	.0933
	(2.16)	(1.83)	(1.26)	(2.63)	(2.19)	(1.83)	(2.79)	(2.06)	(1.87)
Birthweight - male	.294	.385	.223	.078	.0886	.0787	.106	.133	.0769
	(1.65)	(1.48)	(0.91)	(2.09)	(1.84)	(1.38)	(2.66)	(2.08)	(1.58)
t-statistic: difference male - female [p]	-2.23 [.026]	-1.81 [.071]	-1.33 [.185]	-2.45 [.015]	-1.91 [.057]	-1.71 [.089]	-0.95 [.340]	-0.42 [.674]	-1.17 [.245]
F(2, 414): urban = rural [p]	-	0.15 [.858]		-	0.07 [.968]		-	0.36 [.696]	
Number of twins	830	414	416	832	416	416	832	416	416

Asymptotic t-ratios in parentheses.

Family Planning and Human Capital Investment

Among the policy instruments designed to advance development:

Family planning programs implemented to reduce the size of families

A major rationale is that lowering fertility frees up resources in families for human capital investment

One theoretical underpinning is the "Quantity-Quality" (Q-Q) model

The Quantity-Quality Model (Becker and Lewis, 1973)

$$U = U(C, N, Q)$$

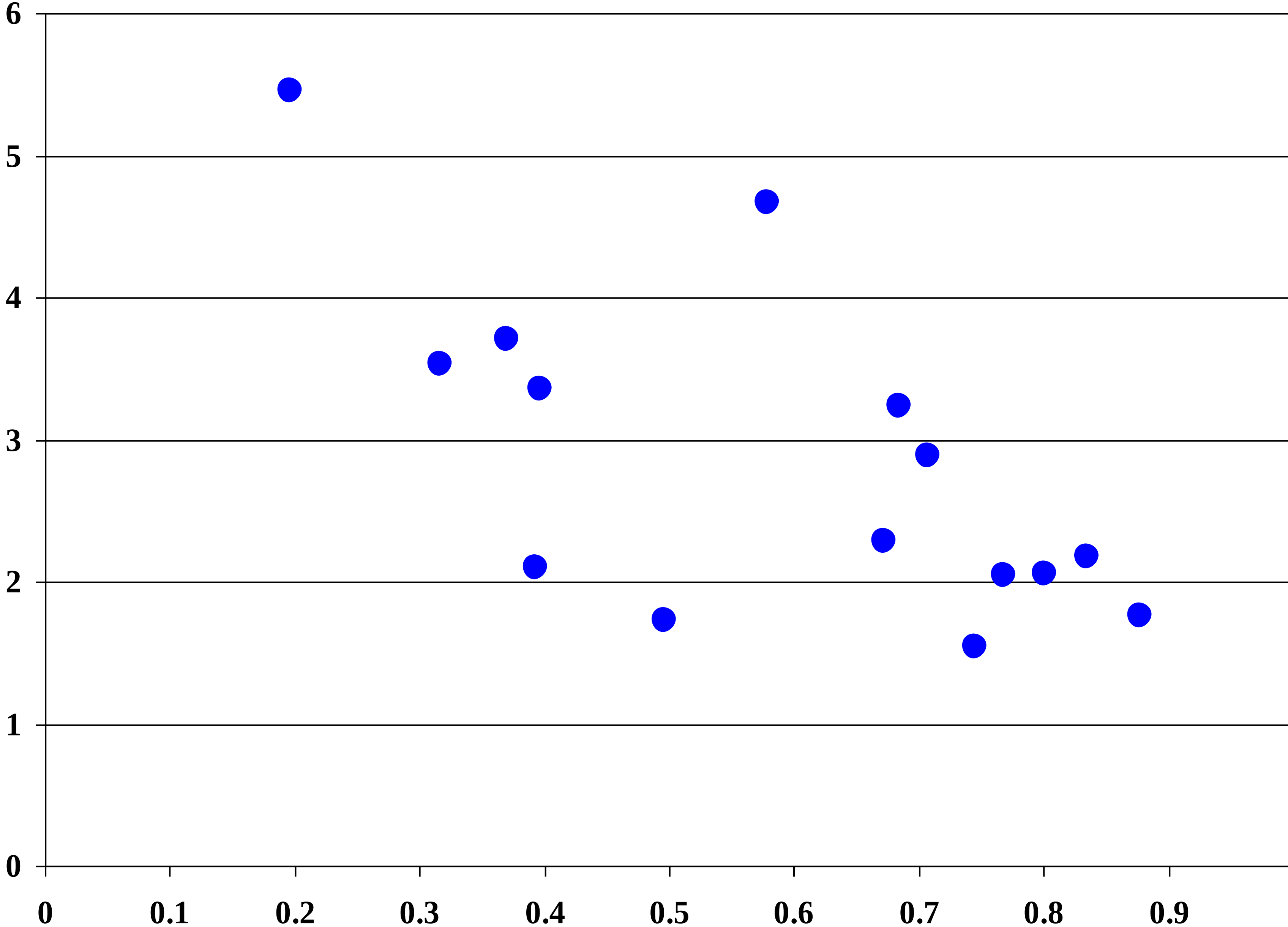
$$Y = C + p_N N + p_Q Q + \pi N Q$$

where N = number of children, Q = *Average* quality per child

$$mrs = U_N/U_Q = (p_N + \pi Q)/(p_Q + \pi N)$$

Can explain why exogenous increases in Y can lead to a reduction in N

Total Fertility Rate and Fraction of 15-19 Year Olds Completing 9 Years of Schooling



Testing the Q-Q model (Rosenzweig and Wolpin, 1980)

$dN/dp_Q < 0$, but what is p_Q? or $dQ/dp_N < 0$, but what is p_N?

Insight from rationing theory (Houthakker and Tobin, 1953):

At the *optimum* levels of C, N, Q

$$dQ/dN = (dN/dp_Q)^C/(dQ/dp_Q)^C = (dQ/dp_N)^C/(dQ/dp_Q)^C$$

If parents allocate resources equally to each child, and quality Q and quantity N are not too complementary, then a decrease in N leads to higher Q

Can use exogenous variation in N to identify sign of *compensated* cross-price effect

But N and Q are endogenous, jointly determined

Twinning is orthogonal to parental preferences and constraints, net of maternal age, so an exogenous determinant of N

Rosenzweig and Wolpin (1980) found support for Q-Q based on India data

Recently researchers again tested the Q-Q model using twins (e.g., Angrist, Lavy and Schlosser (2010), Black, Devereux and Salvanes (2005), Qian (2006), Caceres (2004))

But, the first three studies claim to find no support for the Q-Q model (Israel, Norway, China)

Questions:

1. Do these results reject the Q-Q model? No

2. Is twinning a valid instrument for fertility? No

3. Can twinning be used to understand how reductions in fertility affect human capital investment? Yes

Example: Effects on human capital development of China's one-child policy

How much did this restriction on fertility contribute to growth?

The econometric model they all use, applied to women with *i* pregnancies, is:

$$Q_{i-1} = \beta_n N_i + \mathbf{X}\boldsymbol{\beta} + e_{i-1}$$

$$N_i = \gamma_T T_i + \mathbf{K}\boldsymbol{\gamma} + \varepsilon_i,$$

where N_i = the number of births subsequent to the i-1 birth

T_i = twinning at the ith pregnancy

X, **K** = "vector of control variables"

They find that $\gamma_T > 0$ (but small), and β_n is essentially zero.

That is, the quality of, say, the first child is no different across families who have twins on the second pregnancy and those who do not (and who have less total children)

To obtain a consistent estimate of β_n, T_i must be orthogonal to e_{i-1}

Is β_n really identified using this method?

What is in e_{i-1} (or, relatedly, what is not in $\mathbf{X}$ and $\mathbf{K}$)?

endowments of the individual children

Do these differ across singletons and twins? If so, how does this effect the estimate of β_n?

Endowments and twinning

Recent work shows that endowments matter: in particular, exogenous variations in birthweight significantly affect health and adult outcomes (Behrman and Rosenzweig, 2004; Black *et al*., 2005; Almond *et al*., 2005).

And, twins have substantially lower birthweight than do singleton births (on average 30% lower).

So, twinning not only increases family size, it

(i) lowers the average endowments of children, and

(ii) alters the <u>relative</u> endowments of the children in favor of the first child.

Figure 1. Distribution of Birthweight (Kilograms):
Singleton Births in the NLSY and Twins in the Minnesota Twins Registry

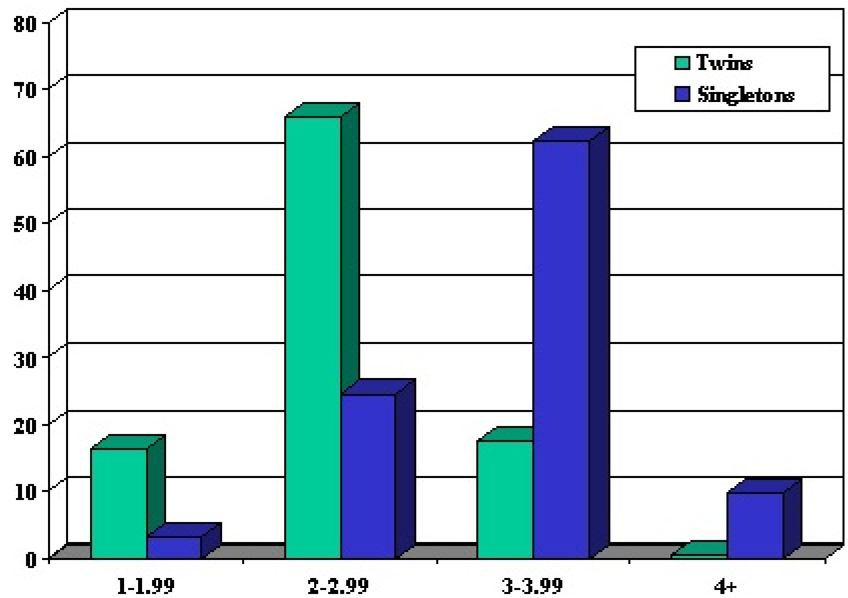
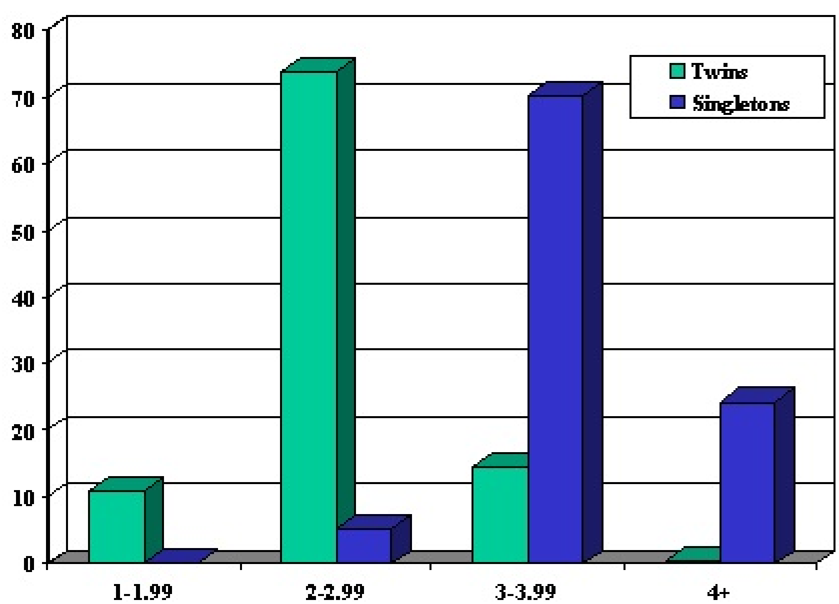


Figure 2. Distribution of Birthweight (Kilograms):
Singleton Births and Twins in Kunming, China



Do changes in the relative endowments of children affect resource allocations across children within the family?

Evidence suggests yes, but little of it (e.g., Behrman *et al*, 1994).

If parents allocate resources across children according to relative endowments then the econometric model employed provides a biased estimate of γ_T - γ_T is not identified: twins are not orthogonal to the error term in the quality equation.

What do we learns from these estimates? How are they biased? How are inter-child resources allocated across children?

To answer these questions formulate two heuristic models and then apply them to data on Chinese twins and non-twins households.

Optimizing family with one child then having a second birth

Model 1: Maximize average child quality Q given a fixed budget for children Y

(1) $$\max Q = (nh_2' + h_1)/(1 + n),$$

where n = number of children at second pregnancy, h_1=quality of child 1, h_2' = average child quality of second birth(s) .

(2) $$h_1 = h(Z_1, e_1), \qquad h_{Z1}>0,\ h_{e1}>0,\ h_{Z1e1}>0$$

where Z_1=resources allocated to child 1, e_1=endowed quality of the first child

Second children (if twins) are identical with endowment e_2, and

(3) $h_2' = h(Z_2', e_2)$, $h_{Z2}'>0$, $h_{e2}'>0$, and $h_{Z2e2}'>0$,

where Z_2'= per-child resources allocated to each of the second-pregnancy children.

The budget constraint is

(4) $nZ_2'p_2 + Z_1p_1 = Y$,

where Y=total income allocated to children, the p_i = the prices of the Z-goods allocated to the children of order i.

Equation (4) is the standard interactive Q-Q budget constraint for the second births (α=1) - with children of order 2 treated equally.

Because twins have lower endowments, it can be shown that:

> The estimate of the second-birth twins effect on the first (second) born’s schooling will be too negative (positive) relative to the appropriate estimate of having an extra child on children’s *average* human capital.

This assumes, as in the simple model, parents allocate more resources to the better-endowed child (because that has a higher payoff by assumption).

If parents dislike inequality across children, and compensate the less-endowed with more investment, then the biases go the other way (still biased).

Need to know allocation rule or control for endowments.

The Data

The Chinese Child Twins Survey (CCTS):

Probability sample of all households with twins aged 7-18.

Probability sample of all households with any children aged 7-18.

Resident in Kunming district (capital of Yunnan Province) in 2002

"One Child" rules in Kunming:

A. One child only in urban areas (non-exempt).

B. Up to two children in rural areas (exempt).

Rules well-enforced and crossing urban-rural essentially prohibited.

Details:

Households with twins identified from the Census by the Urban Survey Unit of the National Bureau of Statistics (same year and month of birth), visited to assure.

Interviewed successfully 1,694 households:

1300 with twins on the first birth 394 twins on the second.

1500 singletons in the same age group.

Collected large array of schooling performance and attainment variables.

One note: many “missing” 16-18 year-olds because many had exited the household, so to avoid selectivity the focus is on twins aged 7-15.

Table 1
Distribution of Fertility for Women Aged 35+, by Exemption Status and Twinning

	Non-exempt, no first-birth twins	Non-exempt, first-birth twins	Exempt, no first-birth twins	Exempt, first-birth twins
One child	94.4%	0	59.0	0
Two children	5.6	99.1	40.3	92.2
Three children	0	0.9	0.7	7.8

Identifying Compensation or Reinforcement

(14) $$Z_{ij} = \beta e_{ij} + \gamma e_{kj} + \mu_j + \zeta_{ij},$$

where
Z_{ij} = educational input for child i in family j,
e_{ij} = child i's birthweight,
e_{kj} = child k's birthweight,
β = own endowment effect,
γ = cross endowment effect (=0 if only one child),
μ_j = unobserved family effect,
ζ_{ij} = unobserved (to parents) child specific endowment.

If parents reinforce endowment differences, then $\beta>0$ and $\gamma<0$; compensation implies $\beta<0$ and $\gamma>0$.

But, $cov(\mu_j, e_{ij}) \neq 0$

(15) $$\Delta Z_{ij} = (\beta - \gamma)\Delta e_{ij} + \Delta\zeta_{ij},$$

where Δ is the cross-sib difference operator.

The within-family estimator identifies only the difference between the own and cross effect $\beta - \gamma$, which is, however, unambiguously positive for reinforcement and negative if there is compensation.

But, $cov(\Delta\zeta_{ij}, \Delta e_{ij}) \neq 0$ with spaced singletons, so use twins!

Table 4
Effects of Birthweight on Parental Schooling Expenditures per Child: Children Aged 7-14

Sample	First-Birth Singletons		First- and Second-Birth Singletons	First-Birth Twins
Estimation Procedure	OLS		Within-Family	Within-Twin
Birthweight (kilograms)	138.9 (2.90)[a]	88.6 (2.01)	59.4 (1.96)	29.4 (1.95)
Age of child	43.1 (3.43)	56.7 (4.79)	45.2 (5.69)	-
Female child	78.2 (1.31)	68.9 (1.23)	2.75 (0.14)	.814 (0.08)
Mother completed high school	-	480.8 (5.32)	-	-
Mother some college	-	773.0 (4.18)	-	-
Father completed high school	-	161.5 (2.39)	-	-
Father some college	-	557.3 (3.93)	-	-
Intercept (First birth)	-242.6 (1.00)	-480.9 (2.13)	-19.6 (0.75)	-8.99 (1.11)
Number of children	1430	1430	612	1882

[a]Robust t-statistics in parentheses.

Estimating Child Quantity Effects Using First-Birth Twins

(16) $$H_{ij} = \eta T_j + \delta e_{ij} + \lambda a_{ij} + \varepsilon_{ij},$$

where $T_j = 1$ if the child is in a household with a first-birth twin pair,
a_{ij} = the age of the mother at child i's birth,
e_{ij} = birthweight of child i
ε_{ij} = error term

Estimating Child Quantity Effects Using Second-Birth Twins

(17) $$H_{ij} = \eta_0 T_j + \eta_1(T_j \text{ x } F_{ij}) + \eta_2 F_{ij} + \delta_1 e_{2j}^* + \delta_2(e_{2j}^* \text{x } F_{ij}) + \lambda a_{ij} + \zeta_{ij},$$

where $T_j = 1$ if the child is in a household with a second-birth twin pair,
$F_{ij} = 1$ if child i is a singleton first-born,
e_{2j}^* = the (average) birthweight of the second-birth children
a_{ij}=the mother's age at the second birth.

Reinforcement implies that $\delta_1>0$ and $\delta_2<0$

Parity-specific twinning effects:

η_0, for second-birth children,

$\eta_0 + \eta_1$, for first birth children, with

$\eta_0<0$ and $\eta_1>0$ implied by the Q-Q model with reinforcement

Table 5
Estimates of *First*-Birth Twinning on Educational Outcomes: Non-exempt Sample (N=1909)

Dependent variable	Expected College Enrollment		Years of Schooling Completed		Math Grade		Literature Grade	
Estimation procedure	Logit		GLS		GLS		GLS	
First-birth twins	-.858	-.685	-.276	-.230	-2.58	-2.12	-1.18	-.944
	(5.61)[a]	(3.65)	(4.85)	(3.34)	(3.50)	(2.27)	(2.04)	(1.37)
Girl	.337	.372	.145	.154	-.132	-.0522	2.32	2.38
	(2.46)	(1.28)	(2.40)	(2.58)	(0.18)	(0.07)	(4.27)	(4.34)
Birthweight	-	.253	-	.0641	-	.652	-	.369
		(1.60)		(1.27)		(0.76)		(0.62)

All specifications include the child's age, age squared and the mother's age at first birth.
[a]Absolute values of t-statistics in parentheses corrected for error clustering at the household level.

Table 7
Estimates of *Second*-Birth Twinning
on Educational Outcomes of Twins and First-birth Non-Twins:
Exempt Sample of Households with Two or More Children (N=1709)

Dependent variable	Expected College Enrollment		Years of Schooling Completed		Math Grade		Literature Grade	
Estimation procedure	Logit		GLS		GLS		GLS	
Second-birth twins	-.658 (4.77)[a]	-.424 (2.21)	-.626 (5.48)	-.649 (5.67)	-3.65 (2.99)	-3.10 (2.44)	-2.78 (2.51)	-2.41 (2.05)
Second-birth twins x first-birth (non-twin)	.490 (2.81)	.327 (1.78)	.319 (1.78)	.424 (2.21)	.672 (0.39)	.548 (0.31)	1.34 (0.98)	.533 (0.37)
First-birth (non-twin)	-.225 (1.74)	1.12 (1.90)	-.274 (2.28)	-1.11 (1.83)	-.615 (0.56)	.176 (0.03)	-.160 (0.17)	6.39 (1.28)
Mean birthweight of second birth	-	.626 (2.96)	-	-.0616 (0.42)	-	1.42 (1.07)	-	.985 (0.77)
Mean birthweight second birth x first-birth (non-twin)	-	-.444 (2.28)	-	.280 (1.34)	-	-.253 (0.14)	-	-2.19 (1.35)

All specifications include the child's age, age squared and sex and the mother's age at second birth.
[a]Absolute values of t-statistics in parentheses corrected for error clustering at the household level.

So, what does all this tell us about the effects of the one-child policy on human capital development in China?

Can provide an upper bound estimate:

(maximum Q-Q trade-off) x (max fertility effect)

Maximum Q-Q trade-off: second-birth twins effects in rural areas, net of birthweight

Fertility effect: Qian (2005): .25, based on rule relaxation in a rural area

McElroy and Yang (2000): .33, based on variation in fines

Table 10
The Q-Q Trade-Off and Upper-Bound Estimates of the Percentage Increase in Human Capital Measures from the One-Child Policy

Outcome	Schooling Progress	Expected Proportion Attending College	Math Grade	Literature Grade	Proportion Good or Excellent Health
Max Q-Q Trade-Off[a]	-13%	-27	-3.8	-2.90	-11
Policy Effect[b]	4.3	8.9	1.3	1.0	3.6

[a]Based on the rural population estimates, second births (Tables 7 and 8).
[b]First row x -0.33, the McElroy-Yang policy effect on family size.

Demand-side Intervention: The Mexican *Progresa* Program

Basic Design: promised transfers over three years that condition on pre-program income only and children enrolled and attending school.

1. Avoids disincentive work on work - post-program income does not affect income transfer.

2. Creates subsidy to child schooling.

Administration:

Identify poor communities (495), identify within poor communities poor (eligible) households based on the1997 Census (67%!).

For evaluation:

A random phase-in: randomly select 314 of 495 to receive the program for first two years.

then remainder receive program in the third year (181 “controls”).

Baseline survey and then post-implementation surveys for two years - a full-scale RCT.

Grants vary by age and sex of the children, who must spend at least 85% of the school year in school.

On average the grants equal about ½ to 2/3 of child wages and 44% of the male adult agricultural wage (635 pesos per month)

Table 1

Monthly Payments for Progresa Program Eligible Families for Children who attend at least 85 Percent of Days[a]

Educational Levels of Students Eligible for Payments		July - December 1998[b]
Primary School - both sexes		
3rd Year		70
4th Year		80
5th Year		105
6th Year		135
Secondary School		
1st Year	Males	200
	Females	210
2nd Year	Males	210
	Females	235
3rd Year	Males	225
	Females	255

Source: Progresa Staff

[a] Excluding those days for which medical or parent excuses were obtained, accumulated over the last two months.

[b] Corresponds to school year first-term, September to December, 1998.

Source: Estimated by the author based on the two pre-program rounds of the survey for only children who are matched in all five rounds or the Panel Sample.

Table 3

Differences Between Enrollment Rates Between Progresa and Non-Progresa Poor Children and Over Time.
(Significance Levels in Parentheses Beneath Differences)[b]

Year of Schooling Completed in Previous Year	Pre-Program Difference of Poor Progresa - Non-Progresa D1			Post-Program Difference of Poor Progresa - Non-Progresa D1			Post-Preprogram Difference in Differences DD1		
	All	**Female**	**Male**	**All**	**Female**	**Male**	**All**	**Female**	**Male**
0	.009 (.351)	.010 (.433)	.007 (.615)	-.002 (.854)	-.010 (.564)	.006 (.742)	-.011 (.482)	-.021 (.353)	-.001 (.969)
1	.001 (.410)	-.009 (.816)	.010 (.376)	.022 (.008)	.007 (.418)	.036 (.002)	.020 (.136)	.016 (.652)	.025 (.070)
2	-.004 (.276)	-.013 (.386)	.006 (.506)	.020 (.009)	.018 (.796)	.021 (.001)	.023 (.226)	.031 (.693)	.015 (.030)
3	.015 (.278)	.025 (.162)	.005 (.882)	.032 (.008)	.013 (.679)	.049 (.001)	.017 (.219)	-.012 (.508)	.044 (.014)
4	.008 (.500	-.016 (.836)	.030 (.266)	.041 (.001)	.038 (.261)	.044 (.001)	.033 (.053)	.055 (.335)	.013 (.064)
5	.015 (.129)	.005 (.544)	.025 (.125)	.047 (.001)	.055 (.232)	.041 (.000)	.032 (.146)	.050 (.647)	.017 (.077)
6	.024 (.345)	.048 (.433)	-.019 (.002)	.111 (.002)	.148 (.001)	.065 (.317)	.087 (.004)	.100 (.070)	.085 (.005)
7	-.012 (.894)	-.005 (.854)	-.015 (.958)	.013 (.147)	.025 (.533)	.003 (.006)	.025 (.378)	.030 (.583)	.018 (.062)
8	-.030 (.913)	-.051 (.932)	-.016 (.836)	.001 (.162)	.015 (.575)	-.010 (.100)	.031 (.347)	.066 (.687)	.006 (.235)
9 or More	.103 (.534)	.327 (.001)	-.156 (.006)	.066 (.317)	.111 (.042)	.026 (.813)	-.037 (.914)	-.216 (.044)	.182 (.020)

Notes: [a] For definition of D1 and DD1, see Figures 1 and 2 and text

Table 7

Cumulative Expected Enrollment Years for Birth Cohort of Poor Children who Enroll and Complete Grade 1

Grade Completed	Preprogram Rounds 1 and 2		Post-Program Rounds 3, 4, and 5		Difference in Differences	
	Progresa	Non-Progresa	Progresa	Non-Progresa	DI	DDI
1	.977	.975	.975	.953	.022	.020
2	.936	.938	.939	.899	.040	.042
3	.896	.884	.904	.837	.067	.041
4	.856	.838	.866	.768	.098	.080
5	.816	.786	.825	.695	.130	.100
6	.464	.428	.511	.352	.159	.121
7	.436	.407	.484	.330	.154	.125
8	.414	.399	.450	.306	.144	.129
Expected Total Years Enrolled for Both Sexes	6.80	6.66	6.95	6.14	.81	.66
Years Enrolled Females	6.66	6.62	6.95	6.19	.76	.72
Years Enrolled Males	6.93	6.72	6.96	6.11	.85	.64

Impact: elasticity of enrollment to school cost = -.2

50% reduction in opportunity costs

10% increase in schooling attainment (.66/6.8)

Earnings?

Assume 10% rate of return on schooling.

Then, using Mincer earnings equation: 0.10*0.66 = 7% increase in earnings.

But, per-capita GDP:

Mexico: $10,000 (2019)   US:$65,000 (2019)

Todd and Wolpin (2006)

The *Progresa* program evaluation ran for a short period - can we do better in anticipating its long run effects?

Is this the best program for increasing schooling? Alternatives?

Estimate a structural model - obtain estimates of fundamental parameters (preferences, technology, constraints) and use them to carry out policy experiments of any type.

But how do we know the model is a good one - what is validation?

T-W estimated a structural model using *Progresa* baseline data.

Assessed if estimated model predicts the program effect.