Webinar Series in Applied Quantitative Analysis - Updated

Date	Topic	
	Session One	
February 29	Potential Outcomes and Omitted Variable Bias I (Theory)	✓
March 7	Potential Outcomes and Omitted Variable Bias II (Application)	
	Session Two	
March 21	Difference-in-differences I (Theory)	✓
March 28	Difference-in-differences II (Application)	
	Session Three	
April 25	Power analysis, clustering and sample size calculations I (Theory)	✓
May 2	Power analysis, clustering and sample size calculations II (Application)	
	Session Four	
May 23	Propensity score matching (Theory)	✓
May 30	Propensity score matching (Application)	
	Session Five	
June 20	Fixed-effects I (Theory)	✓
June 27	Fixed-effects II (Application)	
	Session Six	
July 25	Instrumental variables I (Theory)	✓
August 1	Instrumental Variables II (Application)	
	Session Seven	
August 22	Lagged dependent variables and the Arellano-Bond Estimator I (Theory)	
August 29	Lagged dependent variables and the Arellano-Bond Estimator II (Application)	

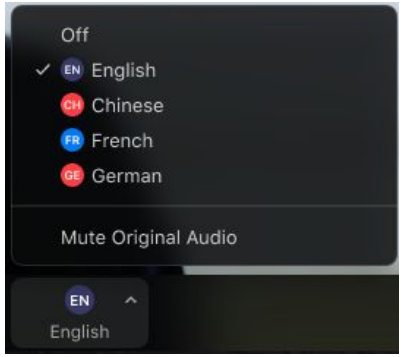
AIR®
Advancing Evidence.
Improving Lives.

https://www.airp3-africa.org/applied-quantitative-analysis-webinar-series

AERC
CREA

Écoute de l'interprétation d'une langue
Windows | macOS

1. Dans les contrôles de votre réunion/webinaire, cliquez sur Interprétation .
2. Cliquez sur la langue que vous souhaitez entendre. (Nous aurons le français) Pas besoin de choisir l'anglais, c'est la langue de la salle Zoom principale



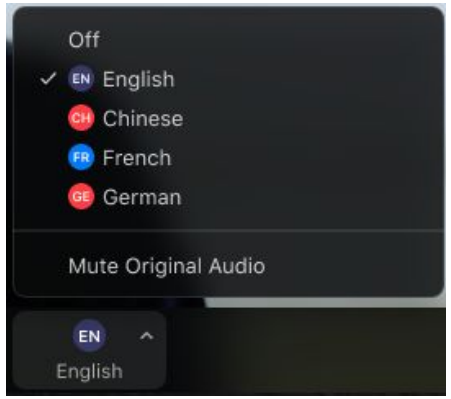
3. (Facultatif) Pour entendre uniquement la langue interprétée, cliquez sur Couper le son original.

Remarques:

- Vous devez rejoindre l'audio de la réunion via l'audio/VoIP de votre ordinateur. Vous ne pouvez pas écouter l'interprétation linguistique si vous utilisez les fonctions audio de connexion ou d'appel téléphonique.
- En tant que participant rejoignant une chaîne linguistique, vous pouvez retransmettre sur le canal audio principal canal si vous réactivez votre audio et parlez.

Listening to language interpretation
Windows | macOS

1. In your meeting/webinar controls, click Interpretation .
2. Click the language that you would like to hear. (We will have French) No need to choose English, that is the language in the main Zoom room



3. (Optional) To hear the interpreted language only, click Mute Original Audio.

Notes:

- You must join the meeting audio through your computer audio/VoIP. You cannot listen to language interpretation if you use the dial-in or call me phone audio features.
- As a participant joining a language channel, you can broadcast back into the main audio channel if you unmute your audio and speak.

Dynamic Panel Data Estimation I - Theory

Ashu Handa

Institute Fellow – AIR

Kenan Eminent Professor of Public Policy – UNC-CH

August 22, 2024

AIR®

Advancing Evidence.
Improving Lives.

AERC
CREA

Why this topic?

- Large number of panel data sets slowly becoming available in Africa
- Panel data supports more sophisticated statistical methods to address endogeneity (e.g. fixed effects, fixed effects with IV—FEIV)
- The Arellano-Bond (1991) and Blundell-Bond (1998) methods provide a solution to a specific problem where there is an endogenous lagged dependent variable

Panel data sets you should have on your computer!

- World Bank Integrated Surveys on Agriculture (LSMS-ISA)
  - Eight countries, national panel data, multiple waves
  - https://www.worldbank.org/en/programs/lsms/initiatives/lsms-ISA

- Ghana Socioeconomic Panel Survey
  - Three waves, national
  - https://dataportal.isser.edu.gh/index.php/catalog/4

- Transfer Project cash transfer evaluation data sets (panels)
  - Kenya, Lesotho, Ghana; Malawi and Zambia coming soon
  - https://transfer.cpc.unc.edu/datasets/

- Kagera (TZ), NIDS (RSA), Young Lives in Ethiopia, etc
  - https://www.younglives.org.uk/

Why would we have a lagged dependent variable?

- Almost all human behavior is conditioned on past choices – what we did in the past influences what we do in the present
  - Panel data lets us build statistical models that include past choices

- More importantly, many <u>outcomes</u> of interest are highly dependent on their past value

- $Y_{it} = f(Y_{it-1}, X)$ where i is the individual/household and t is time

Why would we have a lagged dependent variable?

- Education: child does poorly this year, more likely to do poorly next year
- Nutrition: child is low weight or small this period, more likely to be low weight or small next period
- Physical health: adult has high BMI this year, likely to have high BMI next year
- Mental health, chronic disease, many other examples
- State dependency: the extent to which an outcome or variable depends on its past value

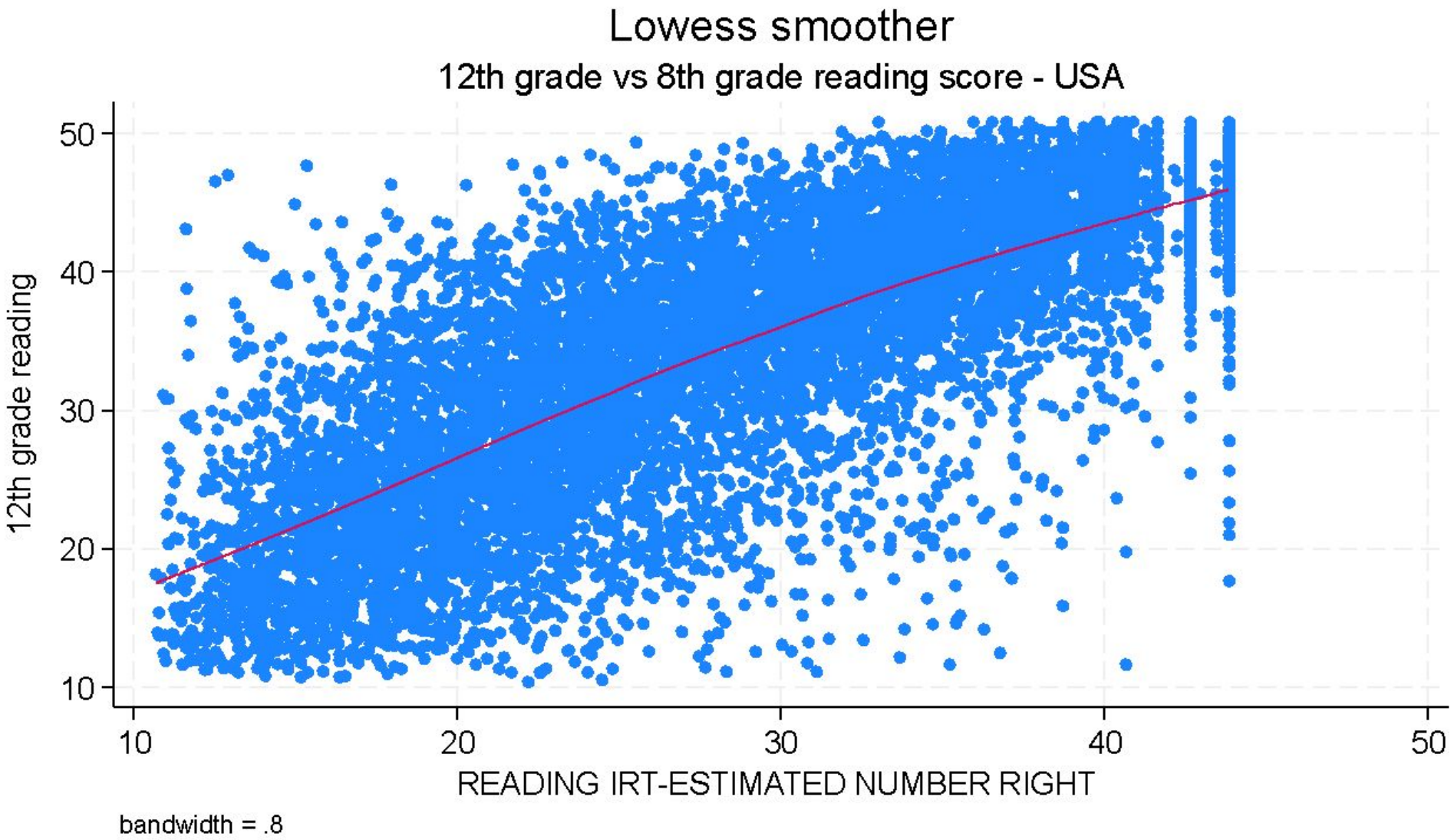
Use the case of child nutrition (height or stunting - H) to illustrate the problem

$$H_{it} = \beta_0 + \beta_1(H_{it-1}) + \beta_2(X_{it}) + \beta_3(X_{ht}) + (\varepsilon_{it} + \mu_i)$$

Current period (t) height depends on lagged height (t-1)

Truly random error: not correlated with Xs nor H_{it-1}

Health endowment of child: fixed over time and correlated with H_i in each period



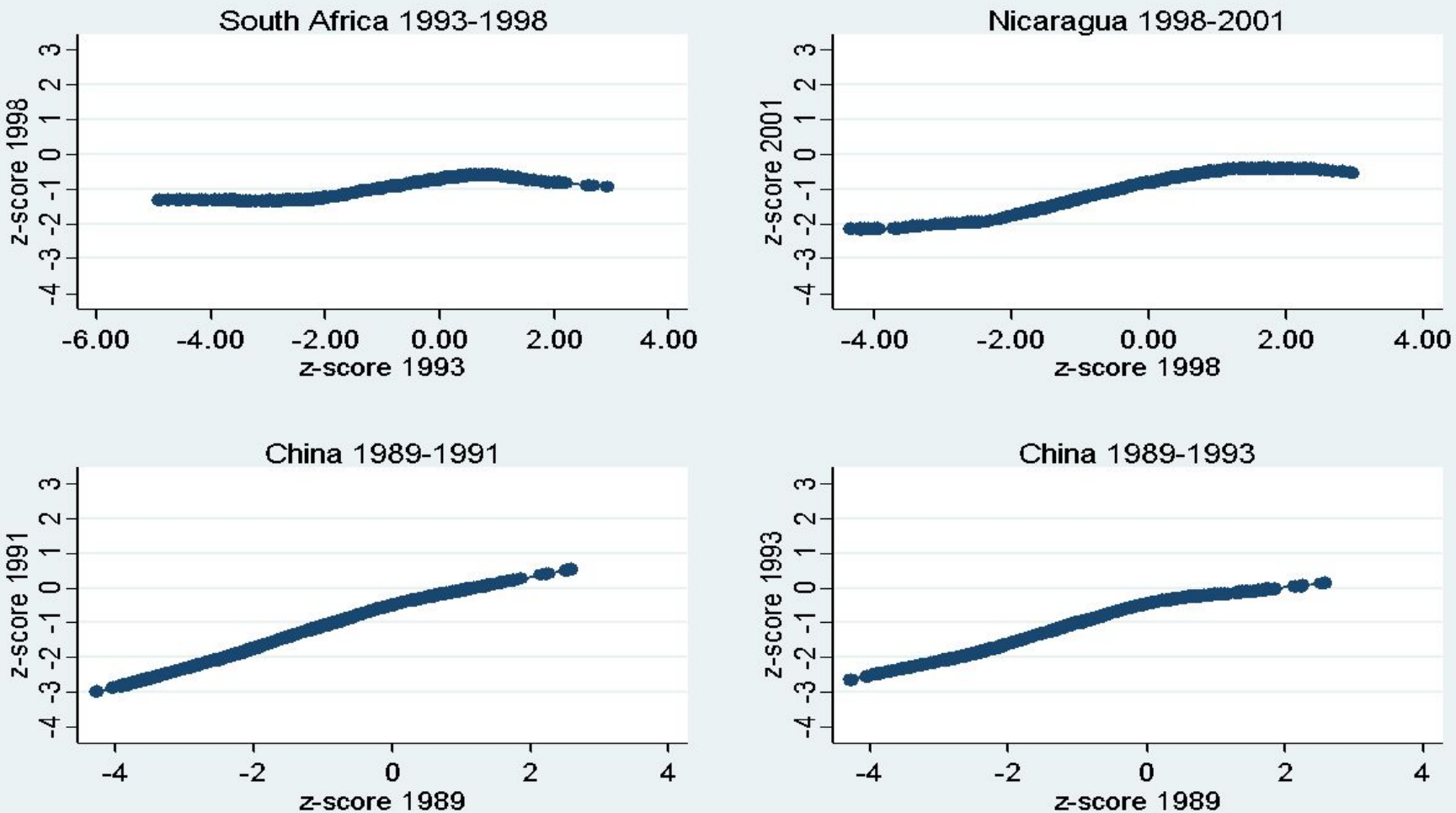
National Educational Longitudinal Survey - NELS

OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 78, 4 (2016) 0305–9049
doi: 10.1111/obes.12117

Is There Catch-Up Growth? Evidence from Three Continents*

SUDHANSHU HANDA†,‡ and AMBER PETERMAN‡

Figure 4: Non-Parametric Relationship of Current and Lagged Height



We have an endogeneity problem: H_{it-1} is correlated with the error term!

$$H_{it} = \beta_0 + \beta_1(H_{it-1}) + \beta_2(X_{it}) + \beta_3(X_{ht}) + (\varepsilon_{it} + \mu_i)$$

We can solve this using IV. Instruments must satisfy two criteria (remember?)

1. Must be highly correlated with H_{it-1};
2. Must not directly affect H_{it};

POLL

Now let us look at the Arellano-Bond (1991) strategy

$$H_{it} = \beta_0 + \beta_1(H_{it-1}) + \beta_2(X_{it}) + \beta_3(X_{ht}) + (\varepsilon_{it} + \mu_i) \quad (1)$$

$$H_{it-1} = \beta_0 + \beta_1(H_{it-2}) + \beta_2(X_{it-1}) + \beta_3(X_{ht-1}) + (\varepsilon_{it-1} + \mu_i) \quad (2)$$

Take the difference in these two equations (t) – (t-1)

$$\Delta H_{it} = \beta_1(\Delta H_{it-1}) + \beta_2(\Delta X_{it}) + \beta_3(\Delta X_{ht}) + (\Delta \varepsilon_{it}) \quad (3)$$

μ_i has been removed in (3) – good
But there is still a problem

Now let us look at the Arellano-Bond (1991) strategy

$$\Delta H_{it} = \beta_1(\Delta H_{it-1}) + \beta_2(\Delta X_{it}) + \beta_3(\Delta X_{ht}) + (\Delta \varepsilon_{it}) \quad (3)$$

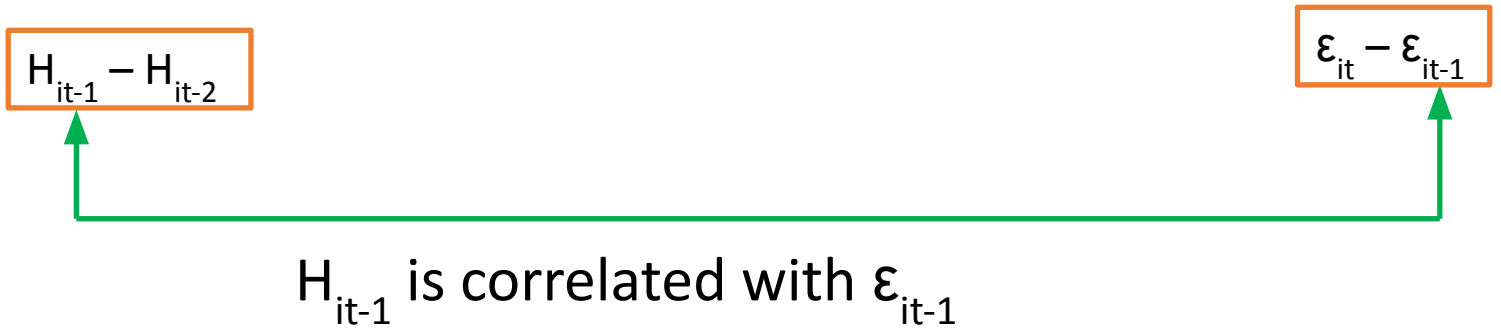
H_{it-1} is obviously correlated with ε_{it-1}

Arellano-Bond propose H_{it-2} as an instrument for $(H_{it-1} - H_{it-2})$
Is this a valid instrument?

H_{it-2} depends on μ_i of course, but u_i is not in (3)!!

Tthe Arellano-Bond (1991) strategy when facing an endogenous lagged dependent variable

$$\Delta H_{it} = \beta_1(\Delta H_{it-1}) + \beta_2(\Delta X_{it}) + \beta_3(\Delta X_{ht}) + (\Delta \varepsilon_{it}) \quad (3)$$



Use H_{it-2} as an instrument for $(H_{it-1} - H_{it-2})$

"Use the two-period value of the dependent variable in levels as an instrument in the differenced equation"

Blundell-Bond 1998 (building on Arellano-Bover 1995)

- When N is large and T is small
  - These are exactly the data sets I listed at the beginning, large, national data sets (N in the thousands) but just a few waves (T around 3, 4 or 5)

- In these data, Y_{it-2} tends to be a very <u>weak instrument</u> for $(Y_{it-1}) – (Y_{it-2})$

- Their solution is a ‘system estimator’
  - XTDPDSYS or XTDPD in STATA (will use it next week)

What is the Blundell-Bond (1998) solution?

$$H_{it} = \beta_0 + \beta_1(H_{it-1}) + \beta_2(X_{it}) + \beta_3(X_{ht}) + (\varepsilon_{it} + \mu_i) \qquad (1)$$

Use $(H_{it-1} - H_{it-2})$ as an instrument for H_{it-1}

But isn't $(H_{it-1} - H_{it-2})$ correlated with μ_i?

H_{it-1} and H_{it-2} are both dependent on μ_i
But when you subtract the two, μ_i cancels out!!

Estimate the levels equation (1) using the lagged difference in Y as the instrument <u>together with</u> the Arellano-Bond (1991) estimator" <u>as a system</u> "system estimator"

Recap of the Blundell-Bond (1998) system estimator

$$H_{it} = \beta_0 + \beta_1(H_{it-1}) + \beta_2(X_{it}) + \beta_3(X_{ht}) + (\varepsilon_{it} + \mu_i) \quad (1)$$

Use $(H_{it-1} - H_{it-2})$ as an instrument for H_{it-1}

Levels equation with a lagged differenced instrument

$$\Delta H_{it} = \beta_1(\Delta H_{it-1}) + \beta_2(\Delta X_{it}) + \beta_3(\Delta X_{ht}) + (\Delta \varepsilon_{it}) \quad (3)$$

Use H_{it-2} as the instrument for $(H_{it-1} - H_{it-2})$

Differenced equation with a lagged levels instrument

TABLE 4

Summary coefficients of lagged height-for-age z-scores on current z-score, China 1991–93

	OLS	*Arellano–Bond*	*Blundell-Bond System*		*2SLS*		
	Levels	*First difference*	*First difference*		*Levels*	*Levels*	*Levels*
Estimator	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Coefficient	0.683	0.091	0.566	0.510	0.530	0.642	0.391
(SE)	(0.023)	(0.110)	(0.064)	(0.085)	(0.070)	(0.083)	(0.084)
t-statistic	29.70	0.83	8.84	6.00	7.57	7.73	4.65
Instrument set							
2 period lagged height-for-age	N/A	Yes	Yes	Yes	Yes	No	Yes
Prices, interactions and wealth	N/A	Yes	Yes	No	Yes	Yes	No
F(instruments)		1.90			11.35	2.1	108.08
Partial R-squared for stage one instruments		0.065			0.296	0.04	0.130

Notes: Observations are 1,121. The first difference estimator uses two period lagged height-for-age as instrument as well as lagged differences in prices and wealth. The levels equation uses the two period lagged difference in height-for-age as the instrument and lagged prices and wealth. See text for details.

Weak instrument problem identified by Blundell-Bond

System estimator

TABLE 4

Summary coefficients of lagged height-for-age z-scores on current z-score, China 1991–93

	OLS	*Arellano–Bond*	*Blundell-Bond System*		*2SLS*		
	Levels	*First difference*	*First difference*		*Levels*	*Levels*	*Levels*
Estimator	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Coefficient	0.683	0.091	0.566	0.510	0.530	0.642	0.391
(SE)	(0.023)	(0.110)	(0.064)	(0.085)	(0.070)	(0.083)	(0.084)
t-statistic	29.70	0.83	8.84	6.00	7.57	7.73	4.65
Instrument set							
2 period lagged height-for-age	N/A	Yes	Yes	Yes	Yes	No	Yes
Prices, interactions and wealth	N/A	Yes	Yes	No	Yes	Yes	No
F(instruments)		1.90			11.35	2.1	108.08
Partial R-squared for stage one instruments		0.065			0.296	0.04	0.130

Notes: Observations are 1,121. The first difference estimator uses two period lagged height-for-age as instrument as well as lagged differences in prices and wealth. The levels equation uses the two period lagged difference in height-for-age as the instrument and lagged prices and wealth. See text for details.

Differenced equation with two period lagged instrument in levels

Levels equation with lagged instrument in differences