Appendices to

China's Economic Growth 1978-2025:
What We Know Today about China's Economic Growth Tomorrow

as forthcoming / published in *World Development*

Carsten A. Holz

1 September 2007 through 30 June 2008: Princeton Institute for International and Regional Studies, 330 Aaron Burr Hall, Princeton University, Princeton, NJ 08544, USA; cholz@princeton.edu

1 July 2008 onwards : Social Science Division, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong; socholz@ust.hk

This appendix discusses data and their sources, reports data, and provides various details omitted from the paper due to space constraints.

Appendix 1. Notes on extrapolations....1
Appendix 2. Data sources and explanations for Figures 1-6 in text....3
Appendix 3. Choice of data in section 4....6
Appendix 4. Data 1978-2003....8
Appendix 5. Labor data....9
Appendix 6. Projections of the Quantity and Quality of Labor, 2000-2025....10
Appendix 7. Production function estimations 1978-2002....11
Appendix 8. Estimation of cointegration equations....15
Appendix 9. Comparison of U.S. vs. Chinese education classification in censuses....19
Appendix 10. Annual secondary and tertiary level graduates in China vs. the U.S.....21

3 July 2007

Appendix 1. Notes on extrapolations

(i) Purchasing power parity factor

The PWT price adjustment factor reflects three separate factors for investment, household consumption, and government consumption. The price adjustment factor in each of these three categories is a geometric mean of two values. (i) One value is an extrapolation from the 1985 PWT 5.6 value. This value in turn appears to incorporate a number of estimates by individual researchers. Ren Rouen and Chen Kai (1994) in a detailed study offer an adjustment factor for 1986 of approximately 3.8963. (I derived this numerical value by multiplying the average annual official exchange rate of 3.4543—the arithmetic mean of the four arithmetic means of beginning-quarter and end-quarter exchange rates in 1994—by Ren Rouen and Chen Kai's convertor for RMB into USD of 1/0.8709. For the exchange rates see *Financial Yearbook 1986*, p. II-39; *1987*, p. II-24.) (ii) The other value is a 1993 Guangdong vs. Hong Kong binary comparison value (downward adjusted in an attempt to reflect a multilateral basis).

(ii) Choice of real growth measure

An alternative procedure for the extrapolations would have been to use real growth rates based on *GDP in international prices*, i.e., to incorporate the price adjustment factor into the past average annual real growth rate. It seems more transparent to operate with output in national prices and then to make explicit assumptions about price adjustment factors for the future. Theoretical support comes from the Gerschenkron effect, namely that, empirically, the use of early-year prices leads to higher estimates of growth rates than the use of late-year prices (explained by the substitution effect from expensive to cheaper products); for developing countries, international prices are like "late-year" prices. Lawrence Klein and Suleyman Ozmucur (2003, p. 21f.), thus, feel that China's purchasing power parity GDP growth rate is likely to be too low. On the other hand, Alan Heston, Daniel Nuxoll, and Robert Summers (1995) find just the reverse to be the case for the aggregate PWT data across the 116 countries in their study; they attribute their findings to the high level of aggregation in the PWT.

(iii)Meaning of coastal data (China)

The average annual per capita real growth rate of China's richer coastal provinces used in the calculations here automatically incorporates the influx of people into the richer coastal provinces throughout the last two decades; i.e., by construction of the average, the newcomers to the rich areas have achieved the same average per capita GDP as the original residents of the rich areas. There is no reason why this should not continue into the future (and the extrapolations here implicitly assume it does), leading to a larger share of China's population (and thus a larger absolute number of Chinese people) enjoying the higher living standard in the future. If the official population data were to cover only the population registered with a regular residence right (*hukou*), the per capita GDP used in the coastal and five-province calculations here should be smaller; this is the case because, in current form, these per capita GDP data then do not incorporate laborers without regular residence in these provinces (which all experience labor inflows), but include the value added they produce. This latter scenario is unlikely since the Chinese population data are based on census data. The 2000 census of 1 November 2000 requires local presence for the past half year to be counted in a particular locality; the 1982 census of 1 July 1982 has a 1-year requirement but migrant labor was probably a phenomenon of negligible size in 1982.

Appendix 2. Data sources and explanations for Figures 1-6 in text

The wage rate is calculated as payments to labor in the national income and product accounts ("labor remuneration") divided by economy-wide employment. Employment values as used in the paper are average annual values throughout, except in the case of the variable "annual absolute change in share of agriculture in employment" (when end-year employment values are used).

The original official employment values of Japan, Korea, and Taiwan (as well as the U.S., used in some ratios) are reported in the source as average annual values.[1] (End-year shares of agriculture in employment are then derived from current and next year average values.) The original official employment values of China are end-year values, and midyear (as a proxy for average) values are then obtained as the arithmetic mean of the relevant two end-year values.

All online resources were accessed on or around 9 February 2004 and on or around 22 December 2006, unless otherwise noted.

<u>Japan</u>

Nominal and real GDP 1960-2002: *World Bank Development Indicators* database (at http://www.worldbank.org/data/). Nominal GDP 1955-59 (following the SNA 68) and 2003-05, and GDP deflator 1956-60 (following the SNA 68) and 2003-05 from the previous Economic Planning Agency: http://www.esri.cao.go.jp/en/sna/menu.html#93sna (accessed 22 Dec. 06). The source of the 2003-05 data also has data for a few earlier years, which do not perfectly match the data from the World Bank source; in the case of labor productivity growth, therefore, the 2002 data from the later source (Economic Planning Agency) were used in deriving 2003 labor productivity growth.

Labor compensation in the income approach to GDP: *Annual Report on National Income Statistics*, various years, for 1953-89; http://portal.stat.go.jp/PubStat/topE.html for 1990-02; 2003-05 from the same source as nominal GDP above.

Employment: Statistics Bureau, Ministry of Internal Affairs and Communications, at http://www.stat.go.jp/english/data/roudou/report/2005/ft/index.htm, accessed 22 Dec. 06. "Agriculture" comprises agriculture (presumably farming and husbandry) plus forestry, but excludes fishery.

Exchange rate: PWT (1960-1970), IFS (1971-2002); Japan Statistical Yearbook 2007 (2003-05).

<u>Korea</u>

All data are from the Korea National Statistical Office (at http://www.nso.go.kr/eng2006/).

<u>Taiwan</u>

Nominal and real GDP through 2000: Government statistics portal (at http://www.stat.gov.tw/bs4/nis/enisd.htm) with online *Statistical Yearbook (Taiwan) 2002*.

[1] This is made explicit only with the Taiwanese employment data. For the other countries, it follows from calculating the average of the monthly values and comparing that to the official annual value.

For labor compensation, in particular, *Statistical Yearbook (Taiwan) 2002*, p. 152; *2003*, p. 152; and various earlier hardcopy issues.

Nominal and real GDP, labor compensation, 2001-05: http://eng.dgbas.gov.tw/public/data/dgbas03/bs2/yearbook_eng/Y094I.pdf, accessed 22 Dec. 2006. The data available in this source for a few earlier years differ slightly from the earlier published values. To avoid the resulting (slight) statistical break between 2000 and 2001, labor productivity growth of 2001 is calculated using both 2000 and 2001 data from this more recent source.

Employment through 2000: online *Statistical Yearbook (Taiwan) 2002*, p. 48; *2003*, pp. 43, 48; various earlier hardcopy issues; 2003 values (needed to derive end-year 2002 employment) from http://www.dgbas.gov.tw/public/data/dgbas03/bs7/yearbook/ch3/3-4.xls#a2, accessed on 20 May 05.

Employment 2001-05: http://eng.dgbas.gov.tw/public/data/dgbas03/bs2/yearbook_eng/y025I.pdf, accessed 22 Dec. 06.

Exchange rate through 2000: online *Statistical Yearbook (Taiwan) 2003*, p. 246; IFS for years prior to 1973; PWT for 1951-1960. (In years for which separate buy and sell rates are given, the sell rate is used.)

Exchange rate 2001-05: http://eng.dgbas.gov.tw/public/data/dgbas03/bs2/yearbook_eng/y139.pdf, accessed 22 Dec. 06.

<u>China</u>

GDP: nominal GDP: *Statistical Yearbook 2006*, p. 57; real GDP growth: *Statistical Yearbook 2006*, p. 59; labor compensation: *GDP 1952-95* (for 1978-92, 1995), various pages (sum across provinces); *Statistical Yearbook*, individual issues with data for 1993, 1994, 1996-2003, and 2005 (as sum across provinces); the share of sum-across-provinces labor compensation in sum-across-provinces income approach GDP is multiplied by the official (production approach) economy-wide GDP figure to obtain an estimate of economy-wide labor compensation. Year 2004 data following the income approach to the calculation of GDP are not available.

Employment: *Statistical Yearbook 2006*, p. 126; 1977 values from *Fifty Years of New China*, p. 2. Between 1989 and 1990, China's economy-wide number of laborers increased by 17.03%, with a 14.12% rise due to a post 1990-census switch to a new time series. Prior to 1990, the published economy-wide number of laborers constituted the sum of laborers across industrial sectors (*hangye*). Since 1990, the economy-wide number of laborers exceeds the sum across industrial sectors significantly in each year, but continues to, as in all reform years, equal the sum across economic sectors (*chanye*; primary, secondary, tertiary). Since the economy-wide number following the new time series for the years since 1990 is the one compiled according to international definitions of employment, the economy-wide number of laborers in the years prior to 1990 was adjusted following the population censuses of 1982 and 1990 (later-year official values rely on population census data). For details see Carsten Holz (2006d). The share of agriculture (taken to be the primary sector) in economy-wide employment is in all years based on the official data; the statistical break 1989/90 occurs across all economic sectors and the total (employment in the three economic sectors always adds up to the total).

Exchange rate: IFS (for 1978-80); *Statistical Yearbook 2001*, p. 586 (for 1981-84); *2006*, p. 734 (for 1985-).

U.S. (all data accessed 22 Dec. 06)

National income and product accounts data are from the BEA website at http://www.bea.gov, with GDP from Table 1.1.5 "Gross Domestic Product," and labor remuneration in form of "Compensation of employees, received" from Table 2.1 "Personal Income and Its Disposition."

Employment data are from the BLS website at http://www.bls.gov/ (unadjusted employment level, series LNU 02000000).

Thoughts on the inclusion of other countries

Other East- or South-East Asian nations could have been included. The Philippines, in particular, would have been an interesting case due to the generally poor economic growth performance (significantly different values of the dependent variable than in the other four countries). But including the Philippines would have meant controlling for religion, and perhaps also for climate and ownership structure. This would imply gaining one degree of freedom (one extra observation, the Philippines), while at the same time losing between one and three degrees of freedom due to the new control variable(s), with no overall gain, if not a loss, of explanatory power.

Singapore or Hong Kong would have required a city-state control variable, and perhaps a second control variable to account for the fact that Hong Kong since 1997 is part of China. Thailand and Malaysia would probably have confirmed the patterns observed for Korea and Taiwan. India may have needed a separate control variable for colonial history. Including further countries would have meant the necessity for more control variables, but the absence of data on one control variable, even if in one country only, would then have brought about the loss of this control variable, causing omitted variable bias.

Appendix 3. Choice of data in section 4

The time series econometrics in section 4 relies on pre-economic census national accounts data. The 2004 economic census led to a retrospective revision of 1978/93-2004 GDP values in 2006, but not to a retrospective revision of income approach data (needed to derive the labor share etc.). For details on the post-economic census benchmark revision of national accounts data see Holz (2007).

Income approach data have always been published at the provincial level only, and only in nominal form. From 1978 through 2003, the shares of each of the four components in income approach GDP change very little. Data for 2004 were not published. In 2005, the share of labor remuneration drops from its 2003 value of 0.50 to 0.41, and the share of operating surplus jumps from 0.20 to 0.30.

In absolute terms, labor remuneration between 2003 and 2005 increased by 21.75%, i.e., comparable to the typical annual increase in GDP in earlier years (and in the retrospectively revised production approach GDP series in all years) of about 10%. But income approach GDP between 2003 and 2005 increased by 45.03%, and operating surplus by 113.63%. A first hunch is that the in the economic census newly found value added, represented in the 2005 but not in the 2003 value of income approach GDP, has simply been attributed to operating surplus.

If one were to subtract the newly found production approach value added of 2004 (of 2300.24b yuan RMB in the *Statistical Yearbook 2006*, p. 57 vs. *2005*, p. 51) from the 2005 operating surplus, the thus adjusted 2005 operating surplus is 29.57% higher than that of 2003 (with sum provincial income approach data calculated from the *Statistical Yearbook 2006*, p. 67, and *2005*, p. 62), which would seem plausible. However, depreciation increased by 36.98% and net taxes on production by 44.19%, where one would expect an increase, following the typical annual GDP growth rate of about 10%, by around 20% over the two years.

It is, thus, not at all clear how the pre-2005 income approach data would have to be retrospectively revised. Given the growth rates of depreciation and net taxes on production, it does not seem legitimate to simply attribute the in 2004 newly found value added to earlier years' operating surplus. With no pre- vs. post-economic census income approach data available for any one year, it is not even possible to ascertain the changes in data compilation in any one year.

What are the consequences for the calculations in the fourth section of the paper if one were to forge ahead, and by assumption simply attribute all newly found value added to operating surplus, i.e., if one were to (i) calculate, for each year, the ratio of post-economic census retrospectively revised production approach GDP to pre-economic census production approach GDP, (ii) augment income approach GDP (derived from provincial data) by this ratio, and (iii) attribute in full the newly found income approach value added to operating surplus? Two approaches to growth accounting are used in the fourth section. In the one approach that assumes a constant labor share, the results would remain unchanged, i.e., there would be no impact at all. In the other, more elaborate approach that uses the income growth identity, the relative weights of the labor share and of the sum of the depreciation and operating surplus shares in the forecasts would change. The impact would be minimal through 2015 because the growth rates of wages and capital in this period are very similar (and their relative weight, thus, does not matter); there would be a small upward adjustment to the 2015-25 growth rates (as faster growing capital gets more weight).

The assumption that all in the economic census newly found value added belongs to operating surplus appears too crude to proceed further. A very minor handicap is the lack of 2004 data (which would have to be assumed to be proportional to, say, 2005 data).

The data on the quantity and quality of labor in the fourth section relies on past population censuses. The 2005 1% population sample survey could have yielded additional information, but these survey data are not available in a tabulation that breaks down laborers by age and education, and is thus not helpful (beyond questions about the accuracy of the 1% population sample survey vs. the population census).

Overall, given these data complications, it seems preferable to retain the data that underlie the fourth section. This also allows a comparison of the growth predictions for 2000-05 based on these data to the actual, by now known growth rate.

Appendix 4. Data 1978-2003

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Gross domestic product			Income shares in gross domestic product				Number of laborers (billion)	Average years of schooling	Education shares			Real capital, b y. RMB,	
	Nom. (b yuan RMB)	Defl. (2000=1)	Off. real grow.	Labor	Depreciation	Net taxes on prod.	Operating surplus			Primary school	Upper middle school	College and above	investm. deflator (2000 pri.)	GDP deflator (2000 pri.)
1978	362.4	0.3174	111.7	0.496560	0.097062	0.128124	0.278245	0.459063	5.874791	0.301122	0.170157	0.008486	2749.608	2354.2723
1979	403.8	0.3293	107.6	0.513799	0.096169	0.122185	0.267847	0.469033	5.783396	0.316217	0.143832	0.008467	3029.126	2534.2218
1980	451.8	0.3432	107.8	0.511476	0.098178	0.121286	0.269060	0.484319	5.760803	0.326867	0.125707	0.008548	3301.316	2720.0263
1981	486.2	0.3497	105.2	0.526832	0.099723	0.119050	0.254395	0.499914	5.788318	0.334703	0.114849	0.008717	3574.954	2981.4397
1982	529.5	0.3502	109.1	0.535664	0.100167	0.116219	0.247949	0.517864	5.837811	0.340401	0.108246	0.009026	3867.181	3308.4062
1983	593.5	0.3545	110.9	0.535442	0.101235	0.115860	0.247462	0.531081	5.900243	0.346449	0.103756	0.009425	4191.373	3626.5069
1984	717.1	0.3718	115.2	0.536801	0.100116	0.117936	0.245148	0.551481	5.972499	0.352423	0.101061	0.009891	4548.309	3875.9280
1985	896.4	0.4041	113.5	0.529008	0.099484	0.120505	0.250982	0.570896	6.058273	0.357315	0.099803	0.010445	5076.077	4210.9495
1986	1020.2	0.4244	108.8	0.528206	0.104881	0.125085	0.241809	0.587218	6.158283	0.360549	0.100347	0.011187	5733.765	4833.6678
1987	1196.3	0.4497	111.6	0.520210	0.107488	0.124799	0.247505	0.604605	6.270596	0.362084	0.102767	0.012160	6417.825	5400.9083
1988	1492.8	0.5033	111.3	0.517224	0.106701	0.130615	0.245455	0.622572	6.468605	0.368652	0.105128	0.014059	7186.390	5928.7643
1989	1690.9	0.5488	104.1	0.515126	0.113077	0.132888	0.238910	0.634099	6.666102	0.373349	0.108808	0.016417	8020.676	6717.1315
1990	1854.8	0.5783	103.8	0.534191	0.116730	0.130573	0.218506	0.650444	6.863468	0.375724	0.113875	0.019301	8851.721	7514.3278
1991	2161.8	0.6021	109.2	0.521607	0.123265	0.132686	0.222446	0.651200	6.953363	0.373345	0.113062	0.020981	9761.820	8570.6868
1992	2663.8	0.6329	114.2	0.500942	0.128676	0.133742	0.236639	0.658215	7.038792	0.371426	0.112448	0.022791	10814.752	10176.8551
1993	3463.4	0.7188	113.5	0.506673	0.116288	0.138373	0.238666	0.664800	7.122996	0.369984	0.111960	0.024744	11966.836	12060.7728
1994	4675.9	0.8384	112.6	0.511986	0.119137	0.136277	0.232600	0.671315	7.205689	0.368746	0.111680	0.026953	13271.268	13435.2360
1995	5847.8	0.9477	110.5	0.528411	0.123477	0.128526	0.219586	0.677600	7.289890	0.367423	0.111722	0.029475	14748.564	14255.3711
1996	6788.5	1.0142	109.6	0.533974	0.128038	0.125712	0.212276	0.685075	7.448646	0.357653	0.115332	0.032891	16471.326	15602.7077
1997	7446.3	1.0267	108.8	0.527937	0.136264	0.131568	0.204232	0.693850	7.596186	0.348787	0.118467	0.036419	18422.663	17713.6154
1998	7834.5	1.0155	107.8	0.531394	0.144735	0.133998	0.189873	0.702285	7.737696	0.340230	0.121446	0.040095	20512.538	20078.0205
1999	8206.8	0.9914	107.1	0.523849	0.150666	0.135394	0.190091	0.710155	7.874455	0.331786	0.124648	0.044000	22731.288	22722.7281
2000	8946.8	1.0000	108.0	0.513819	0.154022	0.141553	0.190605	0.717395	8.005892	0.323572	0.128259	0.048078	25057.903	24928.0360
2001	9731.5	0.9996	107.5	0.514532	0.157159	0.140750	0.187559	0.725550	8.133471	0.315726	0.132589	0.052258	27427.441	27495.6018
2002	10517.2	0.9939	108.3	0.509225	0.156699	0.140426	0.193649	0.733825	8.257059	0.308102	0.137888	0.056451	29922.827	30258.4914
2003	11739.0	1.0119	109.3	0.496246	0.159006	0.142855	0.201894	0.740860	8.380055	0.300270	0.144599	0.060694		

1: Official, nominal revised GDP (*Statistical Yearbook 2004*, p. 53; *Statistical Abstract 2005*, p. 18); 2: GDP deflator based on nominal GDP and real growth as first published (*Statistical Yearbook* series; "first published" relevant for years since 1988); 3: Official real GDP growth rate (*Statistical Yearbook 2004*, p. 55); 4-7: *GDP 1952-95*, *GDP 1996-2002*, recent issues of the *Statistical Yearbook*; 8-12: (i) laborers and education data include military personnel, (ii) education shares denote the share of laborers with this level of education as highest level of education (source: Carsten Holz, 2005); 13: Capital data, deflated using the investment deflator (early version of Carsten Holz, 2006e, data using investment-based scrap rates); 14: Capital data, deflated using the GDP deflator of column 2 (original nominal data from appendix "Fixed Asset Data" to early version of Carsten Holz, 2006e). Nominal wages can be obtained as (columns) 4*1/8.

Appendix 5. Labor data

Labor data cover quantity and quality data. Both need to be constructed. This requires elaborate procedures which are explained in Carsten Holz (2005a). That manuscript explains the choice of data, establishes a correspondence between education categories in different years, and constructs time series on the quantity and quality of labor for the two periods 1978-2003 and 2000-2025 separately. A brief summary follows below.

The derivation of labor data for the years 1978-2003 relies primarily on the three population censuses (1982, 1990, 2000) and two 1% sample surveys (1987, 1995). First, annual total economy-wide employment values are established (quantity of laborers). Second, the age distribution of laborers in each year is derived by interpolating and extrapolating the data from the three censuses. Third, the number of laborers at each age in each year is broken down by education categories, based on data from the three censuses and the two 1% sample surveys. Economy-wide human capital measures are obtained through aggregation of the final year-, age-, and education-specific data on the number of laborers.

The derivation of labor data for the years 2000-2025 starts with the year 2000 population census data. First, the age-specific population of each year is derived by aging the year 2000 population (and by imposing the year 2000 birth rate on the following years). Second, the number of persons at each age in each year is broken down by education categories, which involves (i) the imposition of school enrollment rates on the youngest age groups (with assumptions about the future development of recent trends), (ii) the imposition of adult education rates except for the youngest age groups, and (iii) a linking of the education characteristics of age cohorts across years. Third, the age- and education-specific population is translated into laborers using age- and education-specific employment rates. The economy-wide number of laborers and economy-wide human capital measures are obtained through aggregation of these year-, age-, and education-specific data on the number of laborers.

Appendix 6. Projections of the Quantity and Quality of Labor, 2000-2025

	Number of laborers (billion)	Average years of schooling	Education shares		
			Primary school	Upper middle school	College and above
2000	0.714409	8.012866	0.322023	0.128666	0.047978
2001	0.720456	8.121377	0.316476	0.132001	0.049943
2002	0.726477	8.225262	0.310170	0.135132	0.052176
2003	0.732213	8.328351	0.303293	0.138031	0.054809
2004	0.738277	8.429856	0.295426	0.140944	0.057955
2005	0.744419	8.525540	0.290574	0.144071	0.061807
2006	0.748436	8.630268	0.283859	0.147869	0.066709
2007	0.753146	8.739400	0.276419	0.152366	0.072457
2008	0.759109	8.853824	0.268099	0.157413	0.078968
2009	0.765298	8.971423	0.259523	0.162841	0.086169
2010	0.771226	9.087972	0.250992	0.168101	0.093810
2011	0.775696	9.204913	0.242083	0.172995	0.101810
2012	0.778933	9.319190	0.233075	0.177402	0.109984
2013	0.780788	9.432353	0.223856	0.181429	0.118242
2014	0.780522	9.547423	0.214194	0.185377	0.126681
2015	0.779117	9.663950	0.204020	0.189114	0.135212
2016	0.776259	9.781167	0.193563	0.192779	0.143936
2017	0.771466	9.902992	0.182228	0.196412	0.152955
2018	0.765336	10.026827	0.170430	0.199982	0.162217
2019	0.757166	10.158197	0.157691	0.204050	0.171948
2020	0.748612	10.291379	0.144724	0.208449	0.181854
2021	0.740170	10.422555	0.132106	0.212959	0.191815
2022	0.733109	10.546202	0.120267	0.217183	0.201424
2023	0.728962	10.655325	0.109798	0.220659	0.210190
2024	0.727125	10.749938	0.101092	0.223477	0.218290
2025	0.725704	10.838512	0.093474	0.226045	0.226240

Data are from Carsten Holz (2005). For education shares, the omitted category is lower middle school. Values for 2000-03 differ slightly from those in Appendix 4 due to (by necessity) different procedures used in the construction of the two sets of data. Each set of data is consistent individually.

Appendix 7. Production function estimations 1978-2002

The traditional approach to growth accounting decomposes economic growth into growth of the factor inputs labor, capital, and "everything else" (also "technological progress," or growth in total factor productivity, TFP). The traditional growth accounting equation is

$$\hat{Y}_t = c + b_L \hat{L}_t + b_K \hat{K}_t ,$$

where the hats denote growth rates, Y the value of constant-price output, L the (physical) quantity of laborers, and K the value of constant-price capital, or

$$\ln(Y_t) = \ln(A_O) + c\,t + b_L \ln(L_t) + b_K \ln(K_t) .$$

Estimating the growth accounting equation for China for the period 1978 through 2002 (growth rates starting 1979), using annual data, yields

$$\hat{Y}_t = 6.3609 + 0.2840\,\hat{L}_t + 0.2383\,\hat{K}_t \quad \text{, or}$$

(t-values) (1.1083) (0.4570) (0.4735)

$$\ln(Y_t) = 4.1769 + 0.0452\,t + 0.3859 \ln(L_t) + 0.3975 \ln(K_t) ,$$

(t-values) (1.4822) (1.0857) (1.4655) (1.0551)

with the time variable t equal to one in 1978.[2] Output is official year 2000 GDP with values for other years obtained by applying the latest published official real growth rates; the data on the midyear quantity of labor (including military personnel) and on the value of capital (in year 2000 prices) are explained in Carsten Holz (2005, 2006e).

In both equations, all coefficients are insignificant.[3] The residuals in both equations are normally distributed and homoskedastic, but according to the Durbin-Watson statistic or the Breusch-Godfrey serial correlation LM test positively serially correlated.[4] Correcting for serial correlation does not lead to significant coefficients.

---

2 The growth rate is defined as percentage change. Using first differences of all variables in natural logarithms yields similar regression results. At China's relatively large annual growth rates, the first differences of natural logarithms constitute an inexact approximation of the growth rates. For example, a 10% growth rate of, say, output, and a 15% growth rate of, say, capital, turn into 9.53% and 13.98% if natural logarithms are used. If growth rates are taken over several years, first differences of natural logarithms represent an increasingly inaccurate approximation of growth rates.

3 The results are sensitive to the particular capital measure used. The capital measure used in the reported results throughout this paper is the one identified as the most preferred one in Carsten Holz (2006e) and reported in Appendix 8. With other capital series, the coefficient values change, but remain insignificant.

4 Gregory Chow and Kui-Wai Li (2002) estimate the growth accounting equation (1) in logarithms for the years 1952-98, i.e., including the pre-reform period, but then excluding 1958-69. They correct for serial correlation. Their regression is also run imposing constant returns to scale. The coefficient estimates are significant and plausible in terms of factor shares. Imposing constant returns to scale on the reform period data here also yields significant and plausible results, which, however, disappear once serial correlation is corrected for. Gregory Chow and Kui-Wai Li ignore a statistical break in the labor series between 1989 and 1990. Their

China's National Bureau of Statistics publishes nominal GDP values and real growth rates of GDP (but no explicit GDP deflator). Each year's nominal GDP value is revised later, approximately one year after first published, but the real growth rates are usually, and implausibly, not revised. Using as measure of output in the growth accounting equation the official nominal GDP values deflated by the implicit deflator as first published (which is presumably final) yields no major changes in estimation results.[5] Using as measure of output in the first equation a Tornqvist real growth rate of value added aggregated *across sectors*, with official sectoral real growth rates weighted by sectoral nominal value added shares (means of previous and current year), produces highly similar and also insignificant coefficients of labor and capital.[6]

The growth accounting equation can also be augmented by human capital. One approach is to include a direct measure of educational attainment in the growth accounting equation. For example, Wang Yan and Yao Yudong (2003) construct a human capital variable for the population in form of average years of schooling and enter it with the same weight in the growth accounting equation as labor. A second approach is to classify laborers by certain criteria, such as age and education, and to weight changes in the number of laborers in each category by their relative wages. Alwyn Young (2003), for example, adopts this second approach, and estimates relative wages from a mix of NBS household survey data of the years 1986-92 and Academy of Social Sciences household survey data of 1988 and 1995. This assumes that laborers were paid their marginal product, at a time (prior to 1992/93) when most goods prices had not yet been liberalized and labor markets were virtually non-existent.[7]

---

labor measure, in logs, jumps from 5.5329 to 6.3909; a note reports that adjustments for the statistical break do not change the results. Carsten Holz (2005) addresses (adjusts for) the statistical break in the labor series and constructs economy-wide *midyear* labor values. Carsten Holz (2006e) raises a number of questions about the capital series used by Gregory Chow (1993) and Gregory Chow and Kui-Wai Li (2002).

5 If real GDP is obtained by applying the implicit deflator as first published to the official nominal GDP series, the labor coefficient in the first equation turns negative and halves in the second equation, but all coefficients remain insignificant. The rationale for using the implicit deflator as first published, i.e., as obtainable from nominal GDP data as first published combined with real GDP growth rates as first published, is the following. The National Bureau of Statistics in May 2005 provided final revised nominal GDP data for 2003, and similarly in previous years. When the NBS first published 2003 nominal GDP data in the *Statistical Yearbook 2004* in fall 2004 (apart from earlier estimates), it is highly likely that it had available final price indices for all sectors. Most price indices are compiled monthly and published with little time lag. In sectors where base year prices are used, these apply only to the directly reporting enterprises, on which economic data are immediately available (and need no revisions later); the deflator for the whole sector is derived from the data reported by these enterprises on output in constant and fixed prices. In other words, the likelihood for a need to revise the implicit GDP deflator of 2003 after fall 2004 is near-zero. It is unclear why the NBS does (usually) not revise its published real growth rates; perhaps it is such a highly publicized and political figure that the originally published real growth rate becomes sacrosanct?

6 Alwyn Young (2003, p. 1232, note 17) reports that using the Tornqvist weighted sum of the sectoral real growth rates reduces the official GDP growth rate by 0.2% per year. His data are for 1978-98. He does not specify which sectoral breakdown he uses. I was unable to replicate his findings. Based on the real growth rates of the three economic sectors (primary, secondary, tertiary), using the Tornqvist method, average annual aggregate GDP real growth in the period 1978-98 was 0.05% below the official one (or 0.06% if the 1978 growth rate over 1977 is included); re-composing, in addition, secondary sector growth from that of industry and construction, and tertiary sector growth from that of all tertiary sector subsectors yields a 0.07% (0.09%) divergence.

7 Wang Yan and Yao Yudong (2003, p. 39) do not adopt Alwyn Young's method because they feel that the surveys on which Alwyn Young bases his estimations are not representative samples. James Heckman (2005) argues that "the low private rate of return to education [of, depending on study, around 4% or 7%] does not reflect the true rate of return in the late 1980s or early 1990s. Labor markets were so distorted in China that

Following the first approach and adding a human capital variable in the form of average years of education of laborers to the growth accounting equation yields

$$\begin{array}{lllll} \hat{Y}_t & = 5.0390 + 0.2740\,\hat{L}_t & + 0.4463\,\hat{K}_t & - 0.5829\,\hat{H}_t\,, \\ \text{(t-values)} & (0.8344)\ \ (0.4366) & (0.7772) & (-0.7771) \end{array}$$

where H denotes the average years of schooling across all laborers. Improvements in the average level of education appears to have a negative impact on GDP growth, but none of the coefficients is significant (and similarly if the growth accounting equation in logarithms is used), and the residuals are serially correlated (correcting for which does not lead to major differences in results).

The fact that all coefficients of the growth accounting equation (without or with human capital) are insignificant suggests that one or more assumption underlying the growth accounting equation is violated. The growth accounting equation can be derived following two different concepts (two different sets of assumptions). One is the concept of a production function, using, for example, the Cobb-Douglas production function, $Y_t = A\,L_t^{b_L}\;K_t^{b_K}$. The underlying assumptions are (i) the existence of an economy-wide aggregate production function (i.e., applicability of the Cobb-Douglas production function to economy-wide aggregates) and (ii) the particular functional form, which implies, among others, constant output elasticities (b_L and b_K do not depend on time).[8] From the production function point of view, the implication of insignificant coefficients is that there existed no economy-wide aggregate production function in reform period China and/or output elasticities in China during the reform period were not constant over time.

The second concept from which the growth accounting equation can be derived is the national income accounting concept. By the definition of GDP from the income side, GDP can be written as $Y_t = w_t\,L_t + r_t\,K_t$, where w denotes wages, and r the rental rate of capital (that parameter which makes the product with capital equal to GDP less labor remuneration). Assuming (i) constant factor shares and (ii) constant growth rates of wages and of the rental rate of capital, a few lines of manipulation yield the growth accounting equation.[9] In this

---

wages did not reflect the true marginal contribution of educated labor to the economy." James Heckman estimates a social return "as high as 30% or 40%" (p. 62). This severely questions, if not invalidates Alwyn Young's weighting of education in the derivation of the growth rate of human capital, and subsequently questions the validity of Alwyn Young's measures of the contribution of human capital to output growth.

[8] Two other assumptions are no factor substitution between intermediate inputs and labor and capital, and an elasticity of substitution of unity. Switching to gross output value as the output variable and including intermediate inputs on the right-hand side for the period 1978-2002 yields a highly significant coefficient of intermediate inputs but negative contributions to output growth (or output in logs) of labor and capital (both insignificant when growth accounting is in growth rates, and both significant when in logarithms). Switching to the translog production function (which makes no assumptions about the elasticity of substitution) leads to many significant coefficients but significant negative first-order effects of capital and coefficient values of labor and capital in the hundreds rather than in the expected factor share range. (In contrast, Wu Yanrui, 2004, estimates a frontier production function resembling a translog *with restrictions*, using a *provincial-level* panel data set for the years 1981-97, to obtain plausible coefficient estimates.) In both regressions, adding the average years of schooling of laborers does not change the results; education also comes with a negative coefficient (in the second regression it tends to be negative in all terms in which education appears).

[9] The manipulations are: taking derivatives with respect to time, dividing by GDP, expanding right-hand side terms to be able to simplify some combinations of variables to factor shares, assuming factor shares and growth rates of the wage rate and of the rental rate of capital to be constant, integrating with respect to time, then taking

approach, the growth rates of wages and of the rental rate of capital, with standard deviations of 0.0821 and 0.0443, are not sufficiently stable to yield the growth accounting equation as a tautology from the definition of GDP.[10] Consequently, the coefficients in form of factor shares will not be accurately estimated.

A simplification often imposed in the production function approach is to assume, for the Cobb-Douglas production function, constant returns to scale (b_L=1-b_K) and profit maximization in a competitive economy, so that output elasticities equal (constant) factor shares. The literature that incorporates human capital measures in growth accounting exercises for China indeed does not estimate the growth accounting equation but simply inserts fixed values of factor shares as weights (Wang Yan and Yao Yudong, 2003; Alwyn Young, 2003). The regression results obtained above suggest that this is not permissible.

---

antilogarithms. Also see Jesus Felipe and Carsten Holz (2001). The growth accounting equation as presented also follows from other production functions, such as the CES (if the elasticity of substitution approaches unity), the translog (if the additional terms are insignificant), or the general production function $Y_t = F(K_t, L_t, t)$.

10 See Jesus Felipe and Carsten Holz (2001) for the range of variation, determined in simulations, that yields "good" results of an estimated Cobb-Douglas production function (growth accounting equation). In the case of China in the period 1978 through 2002, the labor share is sufficiently stable with a standard deviation of only 0.0111, but the variations in the growth rates of wages and of the rental rate of capital are too high.

Appendix 8. Estimation of cointegration equations

First, for real wages, the potentially relevant variables were inspected visually for trends, then subjected to the appropriate augmented Dickey-Fuller unit root and Phillips-Perron tests for unit roots in levels and in first and second differences. The real wage rate, as many other variables, is an I(1) process. All tests were conducted for variables in levels as well as in natural logarithms.[11]

Second, in order to determine the lag length for cointegration testing, unrestricted vector autoregression (VAR) models were estimated using the real wage rate and one to three other variables (in approximately three dozen different combinations), starting with four lags and reducing the lags one at a time.[12]

Third, the different combinations of the real wage rate with other variables were tested for cointegration at the lag lengths previously determined; since in all cases at least one series exhibited a clear trend, cointegration testing was conducted with a constant in the cointegating equation as well as in the VAR, and, alternatively with, in addition, a trend in the cointegrating equation. If a significant eigenvector (relying on both the trace test and the maximum eigenvalue test) was found, a vector error correction (VEC) model was estimated. Models with one or more insignificant coefficient in the cointegrating equation were dropped.

The table in the paper that reports cointegration equations reports four equations from four different VEC models. Assume any one VEC model contains three variables y, x, and w; the equations for that one VEC model then reads:

$$\Delta y_t = \theta_{11} + \alpha_1 \ (y_{t-1} - \theta_{12} - \beta_{11} \ x_{t-1} - \beta_{12} \ w_{t-1}) + \sum_{i=1}^{1 or 2} \gamma_{11i} \ \Delta y_{t-i} + \sum_{i=1}^{1 or 2} \gamma_{12i} \Delta x_{t-i} + \sum_{i=1}^{1 or 2} \gamma_{13i} \Delta w_{t-i} + \varepsilon_{t1} \ ,$$

---

[11] Results of unit root tests: probability, in %, of rejecting the hypothesis of "has unit root" in levels, i.e., of is I(1) – in first differences, i.e., I(2) – or in second differences, i.e., I(3). Variable A (real wage per laborer) in levels: augmented Dickey-Fuller test with up to 8 lags (determined by the Schwartz Information Criterion) 65 I(1) – 4 I(2) – 4 I(3), and Phillips-Perron test 92 – 44 – 4, i.e., the augmented Dickey-Fuller test indicates a unit root in levels, while the Phillips-Perron test indicates a unit root in levels and perhaps also in first differences. When the augmented Dickey-Fuller test uses many lags (around 7 or 8), the test was also conducted using 1 lag. Logarithm of real wage per laborer: 26 – 8 – 1, 62 – 49 – 1; real capital (B): 100 – 3 – 17 (at lag length 8, 1 at lag length 1), 100 – 42 – 0; logarithm of real capital : 51 – 3 – 33 (at lag length 8, 0 at lag-length 1), 33 – 42 – 0; logarithm of average years of schooling minus 5 (C): 62 – 0 – 13, 30 – 0 – 4; average years of schooling minus 5, squared: 94 – 13 – 1, 84 – 27 – 1 (D); logarithm of share of laborers with primary school education (E): 73 – 4 – 0, 80 – 4 – 0; share of laborers with upper middle school education (F): 99 – 0 – 1, 4 – 0 – 1; share of laborers with college level education or above (G): 99 – 3 – 1, 100 – 24 – 1; number of laborers (H): 97 – 55 – 0, 92 – 4 – 0.

[12] The likelihood ratio statistics with a small-sample multiplication factor of T-c (T numbers of observations, c numbers of restrictions) as well as the Schwartz Information Criterion were used to determine when to stop reducing the number of lags. The number of observations was held constant across all lag lengths, but because the number of observations is so small (annual data for the years 1978-2002), alternatively, the values of the decision criteria were also calculated for the maximum number of observations at each lag length, and a multiplication factor of T was also used as an alternative with the likelihood ratio statistic. As a result of the different criteria, often more than one lag length appeared plausible. In that case, all relevant lag lengths were taken to the next step of cointegration testing.

$$\Delta x_t = \theta_{21} + \alpha_2 \ (y_{t-1} - \theta_{12} - \beta_{11} \ x_{t-1} - \beta_{12} \ w_{t-1}) + \sum_{i=1}^{1 or 2} \gamma_{21i} \ \Delta y_{t-i} + \sum_{i=1}^{1 or 2} \gamma_{22i} \Delta x_{t-i} + \sum_{i=1}^{1 or 2} \gamma_{23i} \Delta w_{t-i} + \varepsilon_{t2} \ ,$$

and

$$\Delta w_t = \theta_{31} + \alpha_3 \ (y_{t-1} - \theta_{12} - \beta_{11} \ x_{t-1} - \beta_{12} \ w_{t-1}) + \sum_{i=1}^{1 or 2} \gamma_{31i} \ \Delta y_{t-i} + \sum_{i=1}^{1 or 2} \gamma_{32i} \Delta x_{t-i} + \sum_{i=1}^{1 or 2} \gamma_{33i} \Delta w_{t-i} + \varepsilon_{t3}$$

The cointegration equation reported in the table in the paper is identically embedded in all three equations as the part in parentheses following α, the speed of adjustment coefficient.

With the variable of interest (wage or capital, in levels or logarithms) as y_t, and x_t and w_t given as in the table in the paper (in the same order as variables appear in the table), the values of α, with t-values in parenthesis, across the four scenarios and different equations are:

y_t = wage in levels: α_1 = -0.296 (-2.128); α_2 = $-1.56*10^{-6}$ (-0.538); α_3 =$5.26*10^{-7}$ (0.754)
y_t = capital in levels: -0.549 (-2.117); α_2 = 0.000160 (1.309); α_3 = $8.05*10^{-6}$ (3.013)
y_t = wage in logarithms: -0.229 (-3.796); α_2 = 0.091 (1.688); α_3 = -0.046 (-2.607)
y_t = capital in logarithms: -0.597 (-4.925); α_2 = 0.024 (0.139)

The speed of adjustment coefficient α_1 of the variable of interest has the correct sign and size in all 4 scenarios. It is significant at just above (slightly larger than) the 5% significance level in two scenarios, and at the 1% level in the other two scenarios. I.e., each of the four cointegration relationships is plausible.

Except for α_3 in two equations, all the α_2 and α_3 are not significant.[13] This suggests that the various labor variables are weakly exogenous to wage and capital, i.e., explain the real wage and real capital but not vice-versa. Granger causality results, as reported in the table in the paper, point in the same direction for wage and capital in natural logarithms (and also for the wage in levels when a slightly different specification than the reported one is used), as does the 10-period variance decomposition for all four cointegration equations (also reported in the table).

The R^2 values of the individual first-difference equations are all in the 0.63 to 0.96 range.

Apart from the results reported in the table in the paper, many attempted cointegration equations have insignificant coefficients, some cointegration equations have implausible coefficient values (and then always a comparatively poor fit when contrasted to the actual real wage rate series in a chart), and a very few, with significant and plausible coefficients, yield similar results to those reported in the table. The ones reported in the table in the paper have the best fit in a visual comparison of the actual and fitted values.

No standard error of forecast GDP growth is available because GDP growth is obtained as the weighted sum of the growth rates of wage and capital. There is a double problem of not having forecasts of wage and capital *growth*, and of furthermore not knowing the covariance of their growth which would also be needed to establish the variance of GDP growth.

[13] In the equation with capital in levels, α_3 has the wrong sign. The product of α_3 and $ß_{12}$ is small, at 0.112.

Standard errors can be approximated for wage and capital separately. The logic of calculating standard errors is as follows: if a cointegration relationship exists, then one should be able to obtain near-identical results by regressing one of the variables onto contemporaneous level values of the others variables, a constant, and leads and lags of their first differences (James Stock and Mark Watson, 1993).

In an OLS regression, the variance of mean prediction for a vector x_0 with values of the X variables for which one wishes to predict is $\text{var}(\hat{Y}_0 \mid x_0') = \sigma^2 x_0' (X'X)^{-1} x_0$, where σ^2 can be approximated by the variance of the error terms in the OLS regression, X is the matrix of explanatory variables, and x_0 is the vector of forecast labor values (one observation).

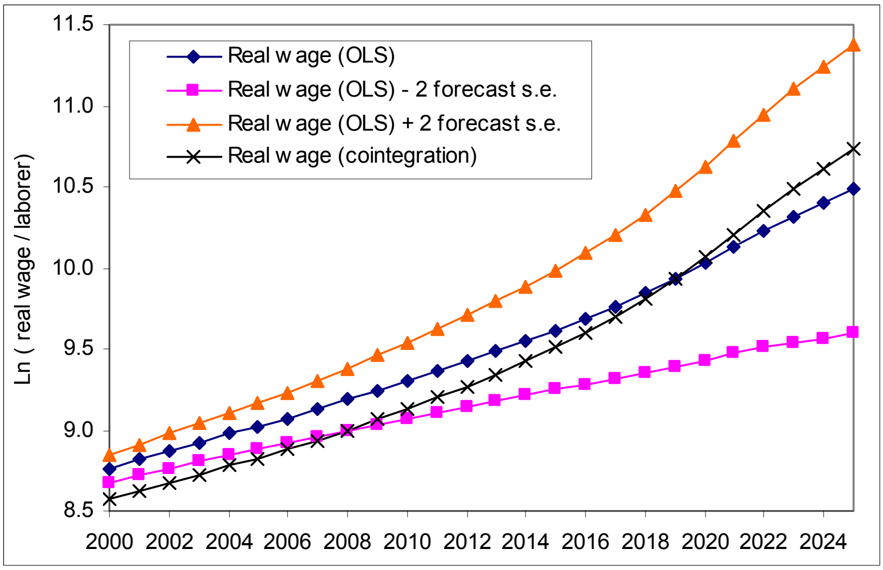
Running an OLS regression of the variables of interest onto contemporaneous level values of the others variables, a constant, and leads and lags of their first differences, the leading and lagged first differences are largely insignificant (and cause the sum squared errors to be very small). Dropping all leads and lags of first differences, i.e., simply estimating the cointegration relationship using OLS (which also increases the number of observations), the values of the coefficients for most variables come close to those estimated in the VEC model (compare the results of Table 1 below to those in the table with cointegration equations in the paper). The sum squared errors of this OLS regression is then used to obtain an approximate variance of mean prediction. Figure 1 and Figure 2 chart a 2-standard error confidence interval for the wage and capital, both in logarithms, and also reproduce the values forecast from the cointegration equation.

For the real wage, the average annual growth rate between 2000 and 2025 of 9.01% (also see table on growth forecasts in the paper) becomes 7.12% in the OLS regression, with a lower and upper value (± 2 s.e. projections) of 3.73% and 10.61%. For real capital, the average annual growth rate between 2000 and 2025 of 10.75% becomes 11.26% in the OLS regression, with a lower and upper value (± 2 s.e. projections) of 10.16% and 12.38%. The confidence interval for capital is smaller than that for the wage, but Table 1 and Figure 2 also show to what extent capital follows a linear trend in 1978-2002 and in the future.

Table 1. OLS Regression of Cointegration Variables

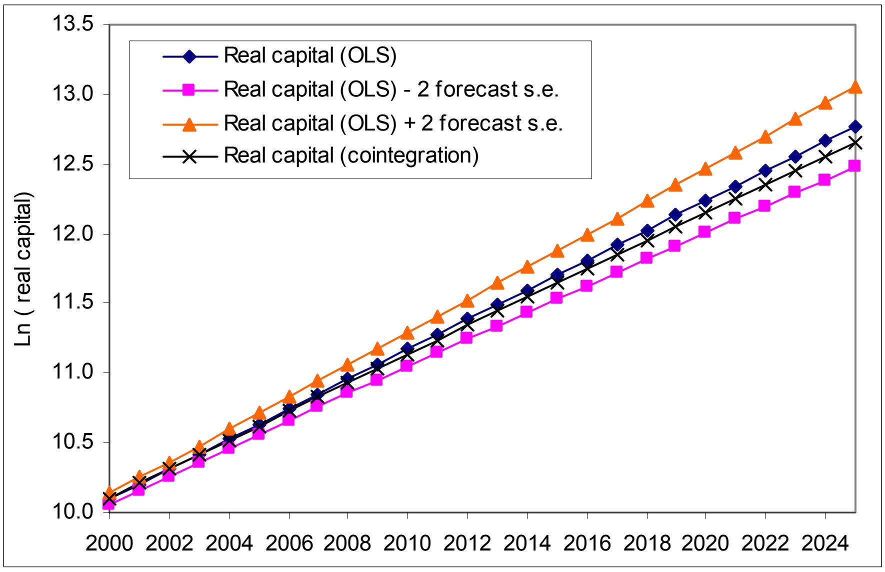
	Ln (wage)	Ln (capital)
Constant	6.6286 (18.371)	7.5853 (137.773)
Average years of schooling minus 5	1.1291 (24.200)	0.1191 (0.954)
Share of laborers with primary school education	-0.7856 (-2.323)	
Trend		0.1036 (12.108)
R^2	0.963	0.996
Sum squared residuals	0.3317	0.0656
Number of observations	26 (1978-2003)	25 (1978-2002)

Values in parentheses are t-values. For variables and their units see notes to table on cointegration equations in paper.



For variables and their units see notes to table on cointegration equations in paper.

Figure 1. Approximate Forecast Confidence Interval for LN(Real Wage)



For variables and their units see notes to table on cointegration equations in paper.

Figure 2. Approximate Forecast Confidence Interval for LN(Real Capital)

Appendix 9. Comparison of U.S. vs. Chinese education classification in censuses

	U.S.	China
	No schooling completed	No schooling
	Nursery school to 4^{th} grade	Literacy class
	5^{th} or 6^{th} grade	
Primary level	[Completion of 6^{th} grade]	Completed primary school (6^{th} grade; *xiaoxue*)
	7^{th} or 8^{th} grade	
	9^{th} grade	
		Completed lower middle-school (9^{th} grade; *chuzhong*)
	10^{th} grade	
	11^{th} grade	
	12^{th} grade, no diploma	
Secondary level	High school graduate	Completed upper middle-school (12^{th} grade; *gaozhong*)
	Some college credit, but less than 1 year	
	1 or more years of college, no degree	
	Associate degree (for ex., AA, AS)	--- Special middle-school (*zhongzhuan*)
Tertiary level (China)		College-level associate degree (*daxue zhuanke*)
Tertiary level	Bachelor's degree (for ex., BA, AB, BS)	Bachelor's degree (*daxue benke*)
	Master's degree (for ex., MA, MS, MEng, MEd, MSW, MBA)	
	Professional degree (for ex., MD, DDS, DVM, LLB, JD)	
		Master's degree (one form of *yanjiusheng*)
	Doctorate degree (for ex., PhD, EdD)	Doctorate (the second form of *yanjiusheng*)

Further details on the U.S.: (i) the associate degree is a degree granted for the successful completion of a sub-baccalaureate program of studies, usually requiring at least 2 years of full-time college-level study; (ii) the bachelor's degree is a degree granted for the successful completion of a baccalaureate program of studies, usually requiring at least 4 years (or the equivalent) of full-time college-level study; (iii) the master's degree is a degree awarded for successful completion of a program generally requiring 1 or 2 years of full-time college-level study beyond the bachelor's degree, (iv) the professional degree is a "first" professional degree which requires at least 2 academic years of work before entrance and a total of at least 6 academic years of work to complete the degree program, including both prior required college work and the professional program itself; depending on the field it corresponds to a Chinese B.A. or possibly a Chinese M.A. or even Ph.D.

Further details on China: (i) upper middle-school comprises three types of schools, the 3-year regular upper middle school (*putong gaoji zhongxue*) which prepares for university entrance, the 3-year vocational upper middle school (*zhiye* [*gaoji*] *zhongxue*), and the 2-3 year technical school (*zhondeng jishu xuexiao*), which in contrast to the first two, is typically run by companies, such as the railway company; a small fourth category of teachers' colleges, *zhongdeng shifan xueyuan*, may be combined with the technical schools under the heading "special middle school," *zhongdeng zhuanye xuexiao*); (ii) there is some ambiguity about the special middle school (*zhongzhuan*): it could be comparable to the upper middle-school, as a 3-4 year program following lower middle-school, or it could be a 2-3 year alternative to college/ university level education following usually upon completion of upper middle-school, and leading to special degrees (such as nursing, in the U.S. an Associate degree) with receipt of a diploma (*biyezheng*, *zige*) rather than an academic degree (*xuewei*) upon completion; it seems that whether *zhongzhuan* denotes a post-upper middle-school education or an alternative upper middle school education depends on the context of the statistics in which the label appears; (iii) the college-level associate degree is typically a three-year program usually duplicating but falling just short of a regular university education (with BA positions strictly regulated and limited by the government, universities but also non-governmental organizations offer this college-level associate degree to those wishing to obtain further education but unable to enter BA programs; this category appears to cover a wide variety of programs, from programs run by universities in parallel to their BA programs, to nursing programs, as in the special middle school, and various programs run by non-governmental or semi-governmental organizations); (iv) medical doctors graduate from a 5-year (rather than 4-year) BA program; (v) the master's degree is a research (thesis) degree of three years' duration (a BA degree is usually a prerequisite); (vi) the doctorate degree is a research degree of three years' duration (a master's degree is usually a prerequisite). Compulsory education in China runs through completed lower middle-school (9^{th} grade).

The table masks three complications in comparability of the two education systems. First, the Chinese special middle-school degree is at least equivalent to the U.S. high school degree, but possibly in some circumstances even to the U.S. post-high school associate degree. Second, while the U.S. and China have comparable BA degrees, the Chinese MA degree differs from the U.S. MA degree in that the Chinese MA degree is a three-year post-BA research (thesis) degree rather than a one- to two-year predominantly taught degree; the Chinese PhD differs from the U.S. PhD in that it requires a prior MA and in that it is usually obtained in three years' time, compared to longer average study times for the U.S. PhD. Third, Chinese professional degrees are usually obtained in form of a BA, but may also take the form of an MA or PhD.

Sources:

U.S.: *Educational Attainment 2000* (possible responses to question of "What is the highest degree or level of school this person has COMPLETED? Mark x ONE box. If currently enrolled, mark the previous grade or highest degree received." in year 2000 census). For the further details in the notes to the table see http://nces.ed.gov//programs/projections/appendix_D.asp#1, accessed on 24 Feb. 2004.

China: *Census 2000*; *Statistical Yearbook 2003*, education section; communication with university professor in China.

Appendix 10. Annual secondary and tertiary level graduates in China vs. the U.S.

	China *(absolute numbers)*		*Ratio: China relative to U.S.*					
			China					
	Regular institutions of higher education (1,000)	Post-graduates (MA, PhD)	high school Def. 1	Def. 2	regular institutions of higher education		research students	
			U.S.					
			High school		Associate, BA	BA	MA, doct., 1st prof.	MA & doctoral
1978	165	9	2.18	2.26	0.12	0.18	0.00	0.00
1979	85	140	2.34	2.40	0.06	0.09	0.00	0.00
1980	147	476	2.02	2.19	0.11	0.16	0.00	0.00
1981	140	11669	1.61	1.84	0.10	0.15	0.03	0.04
1982	457	4058	1.04	1.23	0.33	0.48	0.01	0.01
1983	335	4497	0.81	1.02	0.24	0.35	0.01	0.01
1984	287	2756	0.69	0.92	0.20	0.29	0.01	0.01
1985	316	17004	0.73	1.05	0.22	0.32	0.04	0.05
1986	393	16950	0.85	1.25	0.27	0.40	0.04	0.05
1987	532	27603	0.92	1.41	0.37	0.54	0.07	0.09
1988	553	40838	0.90	1.41	0.39	0.56	0.10	0.12
1989	576	37232	0.89	1.42	0.40	0.57	0.09	0.11
1990	614	35440	0.90	1.50	0.41	0.58	0.08	0.10
1991	614	32537	0.89	1.57	0.39	0.56	0.07	0.09
1992	604	25692	0.91	1.60	0.37	0.53	0.05	0.07
1993	571	28214	0.93	1.64	0.34	0.49	0.06	0.07
1994	637	28047	0.85	1.58	0.37	0.54	0.06	0.07
1995	805	31877	0.80	1.63	0.47	0.69	0.06	0.07
1996	839	39652	0.81	1.77	0.49	0.72	0.08	0.09
1997	829	46539	0.85	1.87	0.48	0.71	0.09	0.10
1998	830	47077	0.93	2.01	0.48	0.70	0.08	0.10
1999	848	54670	0.95	2.07	0.48	0.71	0.10	0.11
2000	950	58767	1.07	2.22	0.53	0.77	0.10	0.12
2001	1036	67809	1.19	2.30	0.57	0.83	0.11	0.13
2002	1337	80841	1.32	2.31	0.71	1.03	0.13	0.15
2003	1877	111091	1.53	2.48	0.95	1.39	0.17	0.20
2004	2391	150777	1.79	2.25	1.16	1.71	0.22	0.25
2005	3068	189728	2.14	2.69	1.47	2.17	0.27	0.31

U.S. degree data of 1978 are for the year 1977/78, and similarly in all other years. Chinese data appear to be based on calendar years. U.S. tertiary sector degree data of 2005 and high school data of 2004 and 2005 are official projections.

The low number of Chinese high school graduates in the mid-1980s could coincide with the reduction in compulsory education from 12 to 9 years, the exact date of which I am not aware of (and which could vary across localities). Conversely, the high number of Chinese high school graduates in the early reform period could be due to upper middle school degrees being awarded for less than twelve years of education.

The first definition of graduates of Chinese secondary schools ("Def. 1") only comprises graduates of regular upper middle schools; the second definition includes graduates of all institutions of secondary education—regular, vocational, and technical schools, and teacher's colleges—except that since 2004 no data are available on the group of technical schools and teacher's colleges (which had 1.484m graduates in 2003 and accounted for 20% of the graduates in China following "Def. 2").

Graduates of Chinese "regular" institutions of higher education (*putong gaodeng xuexiao*) cannot be unambiguously matched with U.S. degrees. Regular institutions of higher education issue BA and college-level associate degrees.

Sources: China: *Statistical Yearbook 1990*, pp. 709, 711; *2003*, pp. 720f; *2006*, p. 800.
U.S.: Department of Education website: http://nces.ed.gov/programs/digest/d05/tables/dt05_101.asp, and http://nces.ed.gov/programs/digest/d05/tables/dt05_246.asp, both accessed on 14 Oct. 2006.

References (beyond those used in the paper)

Annual Report on National Income Statistics. Economic Planning Agency, Government of Japan; various issues.

Chow, Gregory C. "Capital Formation and Economic Growth in China." *The Quarterly Journal of Economics 108*, no. 3 (Aug. 1993): 809-42.

Felipe, Jesus, and Carsten A. Holz. "Why Do Aggregate Production Functions Work? Fisher's Simulations, Shaikh's Identity and Some New Results." *International Review of Applied Economics 15*, no. 3 (2001): 261-85.

Fifty Years of New China. *Xin zhongguo wushi nian tongji ziliao huibian* (Compendium of Statistical Materials for the Fifty Years of New China). Beijing: Zhongguo tongji chubanshe, 1999.

Financial Yearbook. *Zhongguo jinrong nianjian* (China Financial Yearbook). Beijing: Zhongguo jinrong chubanshe, numerous issues.

Heston, Alan, Daniel A. Nuxoll, and Robert Summers. "Comparative Country Performance at Own-Prices or Common International Prices," in M. Dutta (ed.), *Economics, Econometrics and the Link: Essays in Honor of Lawrence R. Klein*. Amsterdam: Elsevier, 1995, pp. 345-61.

Holz, Carsten A. "New Capital Estimates for China." *China Economic Review 17*, no. 2 (2006e): 142-85. (The capital data in section 4 of the paper, i.e., the capital data used in the income growth identity and in production function estimations and reported in Appendix 4, are from the March 2005 version of this paper, with minor differences to the published 2006 ones.)

_____. "China's 2004 Economic Census and 2006 Benchmark Revision of GDP Statistics: More Questions Than Answers?" 2007. Forthcoming in *The China Quarterly*.

Japan Statistical Yearbook 2007. Statistics Bureau & Statistical Research and Training Institute, Ministry of Internal Affairs and Communications. At http://www.stat.go.jp/English/data/nenkan/index.htm, accessed 22 Dec. 06.

Klein, Lawrence R., and Suleyman Ozmucur. "The Estimation of China's Economic Growth Rate." Mimeo, Department of Economics, University of Pennsylvania, 2003.

Ren Rouen and Chen Kai. "An Expenditure-based Bilateral Comparison of Gross Domestic Product between China and the United States." *Review of Income and Wealth 40*, no. 4 (Dec. 1994): 377-94.

Statistical Yearbook (Taiwan). *Statistical Yearbook of the Republic of China*. Directorate-General of Budget, Accounting & Statistics, Executive Yuan, The Republic of China; various years.

Stock, James H., and Mark W. Watson. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica 61*, no. 4 (July 1993): 783-820.