CARNEGIE ENDOWMENT
FOR INTERNATIONAL PEACE

Understanding Energy Intensity Data in China: Insights from a Preliminary Study[1]

Wang Yanjia, Tsinghua University and William Chandler, Transition Energy

Prepared for the Carnegie Endowment
10 December 2010

WHY CHINESE DATA MATTER

This paper attempts to improve understanding of Chinese energy data because doing so is central to understanding global efforts to limit the growth of greenhouse gas emissions. Chinese measures of energy use and economic growth have come under increasing scrutiny since the "Bali Roadmap" emphasized making national emissions trends and reductions for all countries measurable, reportable, and verifiable.[2] In particular, Chinese leaders have claimed success in reaching their unprecedented goal of cutting year 2010 energy intensity—the ratio of energy used per unit of economic output—by 20 percent compared to the base year of 2005. Similarly, China's commitment for year 2020 made at the Copenhagen climate summit in 2009 was to cut carbon emissions intensity by 40-45 percent, again relative to 2005.

Energy production and use accounts for the majority of Chinese greenhouse gas emissions. Unfortunately, the reliability and transparency of Chinese energy data often come into question, particularly in relation to measures of the Chinese economy and the rate of economic growth. Uncertainty stems from problems with and major recent revisions in the Chinese statistical system. For example, a delay until 2010 in publishing Chinese energy data for the year 2008 raised questions among international observers regarding the reliability and believability of Chinese data.[3] Estimates of progress toward the 2010 energy intensity goal have been revised several times, further increasing concern about the data.[4]

---

[1] The authors are grateful to Jeffrey Logan (National Renewable Energy Laboratory), Jonathan Sinton (International Energy Agency), Holly Gwin (Transition Energy), and Howard Gruenspecht (Energy Information Administration) for comments on, and to the Blue Moon Fund and the Rockefeller Brothers Fund for financial support for this paper.

[2] The Bali Roadmap was produced by the Conference of Parties to the United Nations Framework Convention on Climate Change meeting in Bali, Indonesia, 3-15 December 2007. See unfccc.int/meetings/cop_13/items/4049.php.

[3] The National Bureau of Statistics (NBS) of the People's Republic of China usually publishes the China Energy Statistical Yearbook with energy production and consumption data within 12-18 months of the year being reported. For example, the China Energy Statistical Yearbook 2009 would normally have been published by August 2009 to present energy data from 2008, but was published in July 2010. It should be noted, however, that due to publication of the new Economic Census, the NBS had to revise over a decade of energy balance data and had to do so with only a small addition of staff.

[4] Again, it should be noted that multiple revisions in the data are normal. The NBS typically publishes preliminary data in the first quarter of each year and revises it in the second quarter in its Statistical Abstract, and then again makes revisions by October based on some additional detailed data provided by the provinces. The authors would like to thank Jonathan Sinton of the International Energy Agency for his comments on the mechanics of this reporting process.

Chinese officials have recently announced that they will meet their ambitious 2010 target to reduce energy intensity (compared to 2005 levels) by 20 percent.[5] We conclude, with caveats, that those claims are plausible. To explain our view, we describe the sources of the data, how the data were collected, why the data were modified, and make judgments about their quality. We try to show how problems and changes in those data can be better understood and corrected.

HOW CHINESE GDP DATA ARE COLLECTED AND REPORTED

Two Chinese bodies, the National Bureau of Statistics (NBS) and the State Council, lead efforts to collect Chinese gross domestic product (GDP) data, and the NBS provides all of these data in three main publications (see Table 1). The publications are:

- *Statistical Communiqué of the People's Republic of China On National Economic and Social Development*
- *China Statistical Yearbook*
- *Communiqué On the Major Data of the National Economic Census.*

A source of confusion in Chinese GDP data stems from a major new effort to improve GDP data collection. Prior to 2004, the NBS alone collected and published GDP data annually based on surveys sent to firms and agencies. A problem with that approach was that many firms, particularly in the service sector, were omitted from the survey. That was because, by definition, the survey excluded companies with annual revenues below some a particular arbitrary amount. For example, in 2004, firms with revenues less than about 5 million Chinese RMB (US$0.73 million) were not surveyed. This practice significantly skewed the results of the survey, particularly for the service sector.

To improve the data, the Chinese government initiated an economic census to be conducted every four years. That effort was led jointly by the State Council and NBS, with the results published by the NBS. The first census was performed in 2005 and provided more accurate economic data for 2004. The second census was done in 2009, and similarly provided more accurate data for 2008. The census survey is conducted over a period of 12 months. Surveyors are expected to visit every extant economic entity.

GDP since 2005 has been reported in three different versions (see Table 2), though only two versions of GDP may be presented in years in which a census has not been conducted. That is, the NBS normally revises GDP in its Statistical Yearbooks, but not until a current census result is available.

Table 1: China Provides GDP Data in Three Major Publications

PUBLICATION	ISSUED	TIMING	GDP ESTIMATE	NOTES
Statistical Communiqué of the People's Republic of China On National Economic and Social Development	Annual	February	Reported data for first 11 months plus estimated December growth.	GDP in current price
China Statistical Yearbook	Annual	September	Whole-year statistical data	GDP (revised) in current and constant prices.
Communiqué On the Major Data of the National Economic Census	Once in 4 years	January	Census	GDP in current price adjusted between censuses

[5] Xinhua News Agency, "China's "Eleventh Five-Year" emission reduction targets have been completed ahead of schedule" See, the original Chinese language news story at, for example, http://roll.sohu.com/20101201/n300778955.shtml.

Table 2: Published Chinese GDP Vary by Official Publication: An Example from 2007

DATA SOURCE	GDP IN CURRENT PRICE (TRILLION RMB)
Statistical Communiqué of the People's Republic of China On the 2008 National Economic and Social Development	*24.7*
China Statistical Yearbook 2008	*25.7*
Statistical Communiqué of the People's Republic of China On 2008 National Economic and Social Development (After Census)	*26.6*

Sources: On the Major Data of the Second National Economic Census; China Statistical Yearbook 2009.

PROBLEMS ARISING FROM THE GDP DATA

Any revision in GDP will obviously affect estimates of energy intensity, which is expressed simply as a ratio of energy use per unit of GDP. Chinese GDP has been revised substantially as a result of the census estimates, which are considered more thorough and reliable than the NBS GDP estimates. Keeping in mind that GDP is defined as the sum of value-added across all sectors, the largest adjustment to the GDP comes because the census includes more value-added from the service sector (see Table 3). This additional value-added is created mainly by small businesses, which are more likely to be omitted from the NBS data. According to the standard practice of the regular NBS statistical system, value-added data are collected according to a minimum size criterion for organizations, defined mainly by their annual revenue.

Table 3: Chinese GDP Was Revised Based on the 2008 Census (Billion RMB)

	BEFORE CENSUS	AFTER CENSUS	REVISION	
GDP	3,006	3,006	3,140	4.5%
Primary Industry	340	340	337	-0.9%
Secondary Industry	1,461	1,461	1,490	1.9%
Tertiary Industry	1,204	1,204	1,313	9.0%

Sources: On the Major Data of the Second National Economic Census; China Statistical Yearbook *2009.*

In addition to omitting retailers with revenues below 5 million RMB (US$730,000), industrial firms, wholesalers, and other service providers are excluded (see Table 4 for a definition of the designated size of organization for inclusion in data collection in various sectors).

The government does not use a "spot check" method—a test, for example, to see if the aggregate data sum properly—for the services sector, though it does so to cross check industrial enterprises below the designated data collection size. Only income data from enterprises for real estate development, railways, and communication are collected in addition to enterprises in the wholesale, retail, hotel, and catering sectors. Value-added data from many service activities are not included in the NBS surveys, although preparations are being made to do so (see Table 5; all sub-sectors listed in Table 5 are, to date, excluded from the surveys). Most organizations in these sectors are government agencies, but many are non-governmental enterprises that also create value-added.

Table 4: The NBS Annual Economic Surveys Exclude Many Enterprises

SECTOR	CRITERIA FOR INCLUSION IN SURVEY
Industry (Mining, Manufacture, Elec. Gas, Water)	All state-owned and non-state-owned enterprises with annual revenue from principle business ≥5 million RMB
Construction	Various ownership types with qualified certification and independent accounting
Wholesale	Sales ≥20 million RMB Employees (year-end)>20
Retail	Sales ≥5 million RMB Employees (year-end)>60
Hotel	Only hotels with a "star" rating
Catering	Sales ≥2 million RMB Employees (year-end)>40

Table 5: The NBS Annual Economic Surveys Exclude Many Enterprises

SECTOR	NUMBER OF LEGAL ENTITIES	
	Enterprises	Agencies, Institutions, Other144,942
Information Transmission, Computer Service and Software	144,942	8,350
Banking	26,404	1,732
Leasing and Business Services	359,295	67,717
Scientific Research, Technical Service, Geological Prospecting	125,412	76,290
Water Conservancy, Environment and Public Facilities	22,068	35,490
Services to Households and Other Services	106,491	13,977
Education	21,423	313,642
Health, Social Securities and Social Welfare	35,146	46,732
Public Management and Social Organization		1,363,849
TOTAL	**857,122**	**2,118,318**

Source: Communiqué On the Major Data of the Second National Economic Census

The economic census, in contrast, collects value-added data from all organizations. Census takers are instructed to go to every existing enterprise to conduct their surveys. However, even published census data do not aggregate value-added at a sectoral level, leading to substantial differences in estimates (see Table 6). For example, the Statistical Yearbook indicated there were 138,086 wholesale and retail enterprises which created revenue of 18.8 trillion RMB in 2008, while the Census shows the sector had 1,403,000 enterprises and created revenue of 24.7 trillion RMB in the same year.

Table 6: The Economic Census Includes More Economic Activity than the NBS Surveys (2008)

INDICATOR	YEARBOOK	CENSUS	CENSUS/YEARBOOK (RATIO)
Enterprises (units)	138,086	1,403,000	10.2
Employees (persons)	11,374,819	18,912,000	1.7
Revenue (trillion RMB)	18.8	24.7	1.3
Total assets (trillion RMB)	7.5	12.2	1.6

Source: Communiqué On the Major Data Of The Second National Economic Census; China Statistical Yearbook*, 2009*

Before 2009, industrial enterprise accounting statements reported "gross production value" rather than "value-added," a difficult concept to estimate. The concept of value-added in constant prices was even more difficult for firms to estimate. When the Chinese government established its 20 percent energy intensity reduction target in 2006, many firms were confused by the concept and did not know how to measure their mandated energy efficiency improvements. All value-added data for agriculture and industry were estimated by national and local statistical bureaus.

Statistical bureaus mainly adopt the "production approach" combined in some sectors (for example, education or health care) with the "income approach." The production approach estimates the total value of all goods and service produced by all units, minus the total value of input of goods and services of non-fixed assets. This sum, in fact, represents the value-added of all units. The income approach, however, is used sometimes when value-added cannot reliably be estimated. This approach estimates primary income created by all residential units and distributed to residential and non-residential units. The provincial bureaus of statistics (PBSs) independently report provincial GDP to the NBS. Inevitably, there is some double counting of valued-added and so the sum of the provincial GDP is larger than the national GDP. GDP estimated at the provincial level more likely includes double counting because many companies have branches in different locations and conduct business across provincial boundaries.

This inconsistency raises the question as to which estimates are more reliable—the provincial or national data estimates. Some experts think that the NBS data are more reliable because the NBS checks provincial GDP using a growth rate derived from other data such as tax revenues, electricity consumption, and goods turnover. The PBSs, on the other hand, do not make these kinds of cross checks. After adjusting provincial GDP—usually to lower the estimates—the NBS sends the estimates back to the PBSs. The PBSs then publish their provincial GDP numbers using the adjusted data in their provincial statistical yearbooks. Some observers think provincial data are affected by local authorities desire to show higher economic growth rates in their region. The PBSs have never publically responded to questioning of their data quality. At the same time, there has been no reported case of outright data manipulation.

HOW CHINESE ENERGY DATA ARE COLLECTED AND REPORTED

A similar situation pertains to energy statistics. *The Statistical Communiqué of the People's Republic of China on National Economic and Social Development* usually provides figures for total energy consumption and production, coal, crude oil and natural gas production, and total electricity generated. Those data are based on reported information for the first eleven months of each year and a growth rate to estimate values for December, where the growth rate used is the one from the previous December. The *China Energy Statistical Yearbook* later revises the energy data based on a full-year set of statistics. These energy data are then adjusted again after an Economic Census is completed and the revisions are published in the next *Energy Statistical Yearbook*. The *Energy Statistical Yearbook* 2005 revised energy consumption data from 1999 to 2004 based on the results of the first national economic census in 2005 (see Table 7). The biggest adjustment occurred in year 2000 data in which primary energy consumption was revised upward by 6.28 percent compared to the data published initially.

Table 7: Primary Energy Consumption Differs by Source in 1999-2004 (MTCE[6])

	DATA SOURCE		
PRIMARY ENERGY CONSUMPTION BY YEAR	ENERGY STATISTICAL YEARBOOK 2005 (AFTER CENSUS)	ENERGY STATISTICAL YEARBOOK, 2002-2004	DIFFERENCE, COMPARED TO CENSUS (%)
1999	1,283	1,247	2.9
2000	1,324	1,246	6.3
2001	1,357	1,284	5.7
2002	1,441	1,406	2.5
2003	1,672	1,632	2.5
2004	1,939	--	--

The Statistical Communique of the People's Republic of China on 2009 National Economic and Social Development estimated year 2008 energy consumption to be just over 2 percent lower for that year than did the 2008 Economic Census (2.85 and 2.91 billion tons of coal-equivalent, respectively). The Energy Yearbook 2009 also revised upward energy consumption in 2005-2007 based on the results of that second census (see Table 8). Once again, the Energy Statistical Yearbook 2009 revised the energy consumption data from the first national economic census (see Table 9). But the NBS did not explain the reason why the first census data needed to be revised, though the revisions can be striking. For example, the first census had adjusted year 2000 energy consumption by 6.3 percent higher than the original statistical data (see Table 7) and then the second census again adjusted year 2000 energy consumption by an additional 5.3 percent.

Table 8: Primary Energy Consumption Differs by Source in 2005-2008 (MTCE)

	DATA SOURCE		
PRIMARY ENERGY CONSUMPTION BY YEAR	ENERGY STATISTICAL YEARBOOK 2009 (AFTER CENSUS)	ENERGY STATISTICAL YEARBOOK, 2008	DIFFERENCE, COMPARED TO CENSUS (%)
2005	2,257	2,144	5.3
2006	2,475	2,351	5.3
2007	2,684	2,534	5.9
2008	2,775	--	--

[6] MTCE means million tons of coal equivalent, with a ton being defined as the calorific equivalent where one ton equals 29.27 gigajoules.

Table 9: Primary Energy Consumption in 1999-2004 Was Adjusted Upward Based on the Second Census (MTCE)

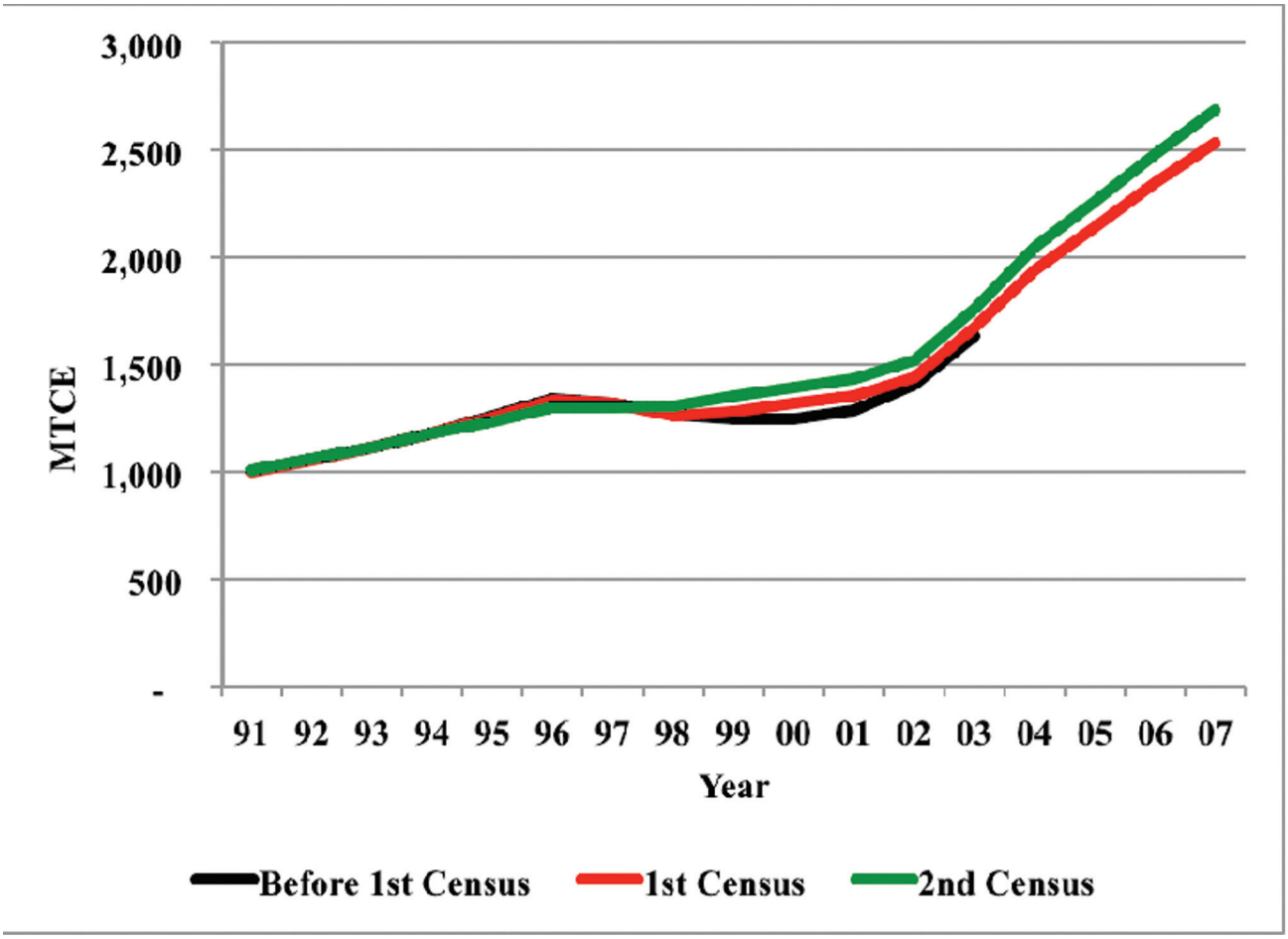
PRIMARY ENERGY CONSUMPTION	ENERGY STATISTICAL YEARBOOK 2009 (SECOND CENSUS)	ENERGY STATISTICAL YEARBOOK 2005 (FIRST CENSUS)	DIFFERENCE AFTER SECOND CENSUS (%)	ENERGY STATISTICAL YEARBOOKS 2002-2004	DIFFERENCE AFTER SECOND CENSUS (%)
1999	1,351	1,284	5.3	1,248	8.3
2000	1,394	1,325	5.3	1,246	11.9
2001	1,430	1,358	5.3	1,284	11.3
2002	1,518	1,442	5.3	1,406	7.9
2003	1,761	1,673	5.3	1,632	7.9
2004	2,042	1,940	5.3		

After the revisions in Chinese energy use data following both the first and second censuses, the supposed decline in energy consumption between 1997 and 2001 was completely eliminated. The original data had shown energy consumption had declined for several years after 1996. Revisions after the first census, which only adjusted data from 1999 to 2003, still reported a decline in national energy use for the years 1997 and 1998. The second census adjusted energy consumption data from 1995, however, and the effect was to lower consumption data for 1995, 1996, and 1997, and to increase the totals after 1998, especially for the years 1998 to 2003. These changes smoothed the curve of energy consumption and consequently removed the decline from the overall trend (see Figure 1).

PROBLEMS ARISING FROM THE ENERGY DATA

The energy-related data revisions brought about by the National Economic Census have mainly resulted in adjusting coal-related data. Coal consumption in 2002, for example, was estimated to be 1.41 billion tons before the census, but was increased to 1.44 billion tons after the first census. There is no adjustment indicated in other energy sources (see Table 10). The data from the second census shows that the statistical quality of coal data collection was not improved after the first census and may have been made worse (compare highlighted data in Tables 10 and 11). That is, coal data were revised even more after the second census than they were after the first census. The statistical data in petroleum, natural gas, and electricity appear more accurate than coal, which is unsurprising since their use is metered. The second census increased petroleum consumption by 0.2 percent and natural gas only 1.5 percent compared to 5.5 percent for coal.

Figure 1: The Two Census Revisions Eliminated the Supposed Decline in Energy Use After 1996



Source: China Energy Statistical Yearbook, *various years*

Table 10: Coal Data Were Revised Most After the First Census (Year 2002 Data)

2002	BEFORE CENSUS	AFTER CENSUS	DIFFERENCE (%)
Primary Energy Consumption (MTCE)	1,406	1,441	2.5
Coal (Million tons)	1,366	1,416	3.7
Petroleum (Million tons)	247	247	0.03
Natural Gas (109 Cubic Meter)	29	22	0.00
Hydro power (109 kWh)	288	288	0.00
Nuclear power (109 kWh)	25	25	0.00

Table 11: Coal Data Were Revised Even More After the Second Economic Census (Year 2007 Data)

SECTOR	BEFORE CENSUS	AFTER CENSUS	DIFFERENCE (%)
Primary Energy Consumption (MTCE)	2,534	2,684	5.9
Coal (Million Tons)	2,586	2,727	5.5
Petroleum (Million Tons)	365	366	0.2
Natural Gas (Billion Cubic Meters)	69.5	70.0	1.5
Hydro Power (Billion kWh)	485	485	0.00
Nuclear power (Billion kWh)	62	62	0.00

The PBSs have no obligation to collect energy data and it is the NBS which collects energy data directly from industrial enterprises. Large energy production enterprises must report their data to the NBS once a month and small energy production enterprises report once a quarter. The NBS checks year-data quality once per year. The NBS issues quarterly data by industrial firm size category, but this is only for internal government or agency use. Only annual data for all industries are published for public use. Energy data are collected on the basis of enterprise sales for supply data and crosschecked with energy production data.

The NBS collects energy consumption data through reporting and spot checks (see Table 12) and makes available national and provincial energy balance sheets based on the data it has. Spot checks in this context means samples taken in proportion to the population size—for example one in 10,000 residences.

Table 12: The Responsibility for Energy Use Data Collection Varies by Sector and Scale

SECTOR	RESPONSIBILITY AND METHOD FOR DATA COLLECTION[4]
Agriculture	Any enterprise accounting for more than 10 percent of provincial agricultural output
Construction	Census
Industry By enterprise above a designated size By enterprise below a designated size	 Reported once a quarter Spot check
Railway, aviation, pipeline	Reported by sectoral administration such as Ministry of Railway, Civil Aviation Administration
Road, waterway, city	Spot check
Catering Services By enterprise above a designated size By enterprise below a designated size	 Report once a quarter Spot check
Residential	Spot check

It is well known that in China it is very difficult to get accurate data on coal. In part, the problem is one of scale as China has more than 1,100 counties with operating coal mines. The comparison of energy balance sheets before and after one of the two census results indicates that most adjustments are in coal data, such as raw coal production, cleaned coal for transformation, and coal consumption by sector (see Table 13). The priorities for statistical improvements thus would appear to be in industry, wholesale, urban residential uses, and the "other" sector.[7]

Table 13: Coal Production and Use Data Were Adjusted by the Census (Million Tons, Year 2007)

ITEM	DATA BEFORE CENSUS	ADJUSTMENT BY CENSUS	CHANGE(%)
Coal output	2,691	165	6.2
Final Consumption of Farming, Forestry, Animal Husbandry, Fisheries and Water Conservancy	15	-8.2	-53.9
Final Consumption of Industry	631	129	20.5
Final Consumption of Construction	6.2	0.5	8.1
Final Consumption of Transport, Storage, Postal and Telecommunications Services	7.4	0.5	6.9
Final Consumption of Wholesale, Retail Trade and Catering Services	18.7	10	53.5
Final Consumption of Urban Residences	29.1	6.3	21.7
Final Consumption of Rural Residences	68.6	10.3	15.0
Final Consumption, "Other"	20.4	12.3	60.3

Source: China Energy Statistical Yearbook *2009, 2008*

Consumption totals among sectors are also revised significantly (see Table 14 for a presentation of final coal consumption, with changes among sectors as a result of the second census). Although switching consumption does not affect estimates for the energy intensity of GDP, switching data may provide information on how to verify the data. For example, fuel oil consumption is shifted from industry to transport. One possible explanation for this shift could be changes in estimates of energy consumption for shipping services.

[7] We assume the differences apparent in the farming sector and for rural residents are caused by a change in data collection methods from measuring coal production to measuring daily coal consumption.

Table 14: A Comparison of Final Energy Consumption Before and After the Census Shows Varying Differences Sector Totals (2007, MTCE)

ITEM	RAW COAL	COKE	OTHER GAS	GASOLINE	DEISEL	FUEL OIL	NATURAL GAS	HEAT	ELECTRICITY
Total Final Consumption	120	-11.4	31.8	0	0.05	1.1	1.3	0	0
Farming, Forestry, Animal Husbandry, Fishery and Water Conservancy	-5.8	-0.24	0	-1.09	-9.56	0	0	-1	-123
Industry	97.2	-11.2	31.1	-0.7	3.6	-4.4	-4.0	-0.34	-4.2
Non-Energy Use	6.5	-0.5	0	-0.02	0.07	-0.21	-1.00	0	0
Construction	0.4	0	0	-0.29	0	0.21	0	0	0
Transport, Storage, Postal and Telecommunications Services	0.40	0	0	-2.21	5.68	5.29	3.99	0	0
Wholesale, Retail Trade, Catering Service	7.36	0	0	-3.24	-6.84	0	0	0	0
Urban Residential Consumption	4.47	0	0.72	3.00	3.50	0	1.33	0	3.20
Rural Residential Consumption	7.59	0.05	0	2.06	1.46	0	0	0	2.21
Other	8.81	0.01	0	2.50	2.19	0	0	0.35	0

Source: China Energy Statistical Yearbook *2009, 2008*

JUDGMENTS REGARDING DATA RELIABILITY

The published official Chinese energy consumption data for natural gas, hydropower, and nuclear power appear to be reliable, at least based on the fact that repeated official reviews of the data have not required major revisions. That is not the case, however, for the coal data, which is unfortunate since coal production and use dominate all important aggregate Chinese energy data including total energy consumption, energy intensity of the GDP, and carbon dioxide emissions. Problems with coal data stem from two characteristics of the fuel: measurements based on the physical quantity of coal consumed, and the measurements of the heat and carbon content of some 2-3 billion tons of coal.

Several sources of information permit estimation of coal quality (see Tables 15a and 15b). This information may be useful for verifying emissions data at some level. Detailed information on coal quality by coal type is normally not available to the public (see Table 16 for common indicators used for coal quality). Carbon content (or fixed carbon content) is not a common indicator collected for coal quality. Anthracite coal from different regions, for example, has different carbon and hydrogen content. Without knowing the carbon content data, it is not easy to calculate the carbon dioxide emissions by type of coal burned. Moreover, the combustion efficiency will differ by coal quality and that in turn will affect coal consumption and carbon data.

Table 15A: Potential Sources of Cross-references for Chinese Coal Quality Data

SOURCE	CHARACTERISTICS INDICATED	DATA AVAILABILITY
Coal Network http://www.coal.com.cn	Output by coal type and region (monthly)	Membership only
China Coal Industry Yearbook 2008 (Census Year)	Output by coal type (anthracite; meager-lean; lean; coking; fat; 1/3 coking; gas-fat; gas; other) and region (yearly)	Public
China Coal Industry Yearbook (Non Census Year)	Output by mine ownership (yearly)	Public
Coal Network http://www.coal.com.cn	Coal quality indicators for specific producers or mines	Public

Table 15B: Potential Sources of Cross-references for Chinese Coal Quality Data

TESTING BASED	AIR DRIED	DRIED	DRY ASH FREE
Inherent Moisture, % Ash, % Volatile Matter, % Heat Value, kcal/kg	Common Coal Quality Indicators		
Fixed Carbon, % Total Sulfur, % Carbon, % Hydrogen, % Nitrogen, %	From Coal Quality Analysis Report Only		

No cost-effective approach has been identified for checking coal quality data after the data have been submitted. The quality of such data can be ensured only during the entire submission process, from bottom to top. Many local officials lack the human or technical capacity to ensure data quality. Research and training are needed to suggest how the data bureaus can ensure common sense is used by officials to judge data reliability. For example, when observers questioned coal data in the year 2001 (see Figure 2), most observers simply assumed that small mines were hiding their output because their production had been declared illegal under government policy, or, alternatively, that firms were hiding production simply to avoid paying taxes. Only a few observers realized that the local bureau of statistics were simply not counting small coal mine production because those officials simply assumed the small coal mines had stopped operating, as supposedly required by the government, and therefore never checked on them to verify that the mines had in fact stopped production. The officials thus did not count small mine production because they did not collect the data. This situation bolstered the belief that local governments were hiding coal production data because the officials were motivated instead to showcase their achievements in increasing coal output, as the government had previously ordered. But such explanations can only explain the reduction in coal output, not in coal consumption data.

There are indicators for judging the reliability of data related to companies and small mines (see Table 16). For example, the capacity of mines owned by Town and Village Enterprises (TVEs) in the early part of this decade increased significantly while their production capacities shrank. It would be easier to draw conclusions about those counterintuitive trends had there been an effort to crosscheck and verify this data. That is, it would be useful to impose a kind of "laugh test" on the publication of data.

Figure 2: TVE Coal Output Suggests a Means of Spot-Checking Coal Data[8]

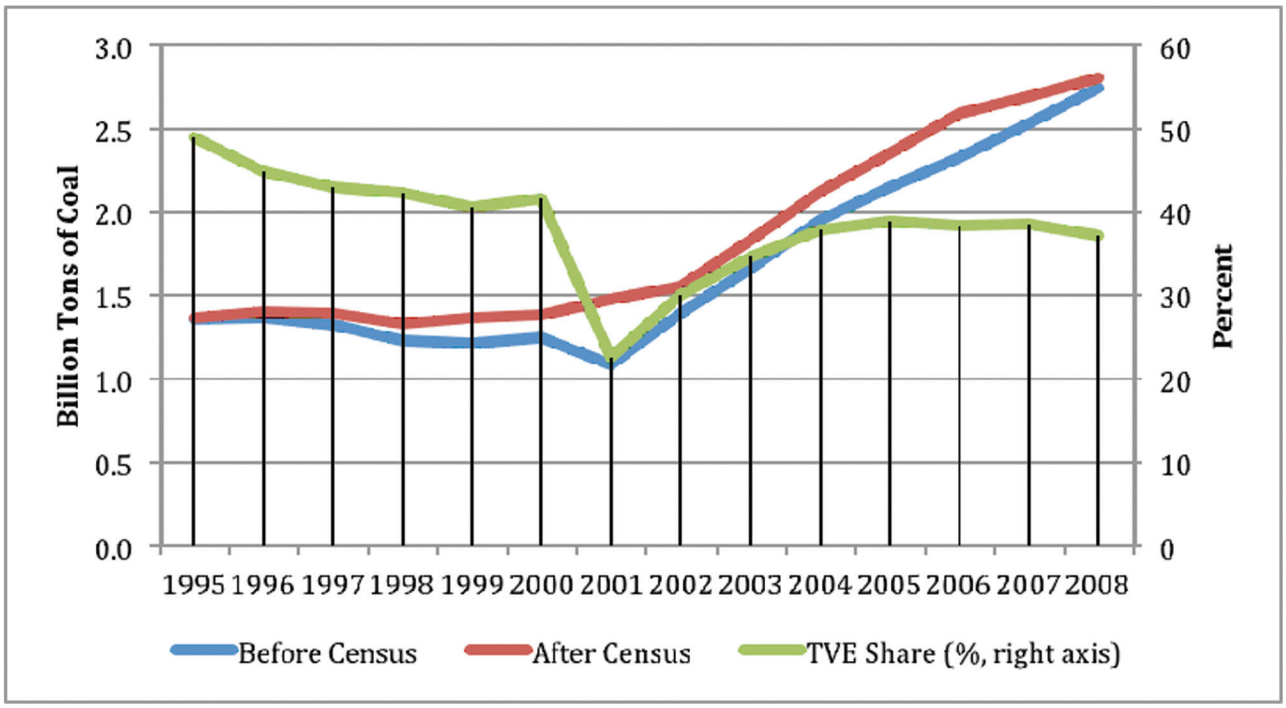


Table 16: Coal Output in Shannxi Province

COMPANY BY OWNERSHIP	ITEM	2007	2008	JUDGEMENT
Shenhua Company	Mines Capacity (million tons) Output (million tons)	6 60 60	6 65 65	High reliability
State-owned Key	Mines Capacity (million tons) Output (million tons)	26 29 30.8	27 29 43.7	Modest reliability
State-owned Local	Mines Capacity (million tons) Output (million tons)	51 18.5 22.6	59 17.9 25.5	Low reliability
TVEs	Mines Capacity (million tons) Output (million tons)	651 44.4 58.5	362 26.6 83.1	Very low reliability

Data source: China Coal Industry Yearbook *2007, 2008. "Capacity" includes mines under construction, but "Output" does not.*

[8] Town and village owned enterprises

It should be feasible to develop software to check such data discrepancies and therefore improve data quality and avoid mistakes—simply to be able to correct data-entry errors. For example, raw coal consumed as feedstock in 2008 was 37.6 million tons, according to the national energy balance sheet. But simply summing the data across all provincial energy balance sheets results in a total of 50.4 million tons. That is, the sum of provincial reports of tons of coal used as a feedstock is one-third higher than the national total. This kind of problem could be avoided by automated cross-checking for data consistency.

Other published data and expert knowledge could be applied to judge provincial data reliability, especially for priority areas such as ammonia output and other feedstocks (natural gas, naphtha) for ammonia production. The data could be crosschecked, for example, by comparison with the regulatory standard established for feedstock use in ammonia production. (See Table 17 for examples of approaches to crosschecking coal consumption data for ammonia production.)

Table 17: Crosschecks Can Be Applied to Coal Consumption Data for Ammonia Production (2008, million tons)

SELECTED PROVINCE	RAW COAL AS FEEDSTOCK	AMMONIA OUTPUT	COAL/ AMMONIA	COMMENT
National	37.6	50.0	0.75	Assumes 5360 kcal/ton raw coal
Provincial Sum	50.4	50.0	1.01	
Shandong	0	6.0	0	Unclear if other feedstock used
Guangdong	4.0	0.1	53.9	Unclear if other products use coal as feedstock
Sichuan	2.6	4.1	0.6	Most of ammonia plants in Sichuan use natural gas as feedstock

FINDINGS AND RECOMMENDATIONS

China publishes same-year GDP and energy consumption data using different values on different occasions and in different publications. When these documents are translated into English or reported by the media, only the numbers themselves receive attention and the qualifiers, or footnotes, sometimes get ignored. And sometimes the data revisions are not fully explained in the Chinese language reports.

Chinese data collection agencies could improve the quality of statistical reporting by making better use of data spot checking, data sampling, and personnel training. The most important thing Chinese leaders could do in response to this situation is to facilitate capacity building at all levels of data management and administration to guarantee data quality (and consistency) from top to bottom. For example, when enterprises set up their energy management systems, a national standard consistent with NBS' requirements on data collection could be followed. When data are maintained in electronic format, the reported information could be transferred directly and automatically and thus minimize human error. For measures of electricity, natural gas, and heat use which are metered and invoiced by utilities, data could be utilized directly for statistical reporting in place of survey data obtained from enterprises.

More effort could be made to collect data on coal output, especially from small mines. For example, small coal mines could be required to report their production data even if their output is quite small.

Significant improvements could be derived by applying lessons-learned from the Economic Census. The statistical shortcomings revealed by the census include those large adjustments required for total coal production and in GDP for the service sector. The Chinese government could concentrate its efforts on improving data collection, analysis, and processing in these areas.

Improving data quality is a long-term task, with no overnight solution. Bringing additional energy expertise to the analysis of energy data systems would likely improve the quality of the official energy data systems. Verification and checking of data quality is a laborious task, but automated crosschecking both minimizes this burden and improves data quality. The national energy balance sheet, obviously, should sum to the total of corresponding items in all provincial energy balance sheets and that it currently does not is indicative of the challenge ahead.[9] The Chinese data system could improve the quality of its products and enhance their credibility by inviting independent organizations and experts to review the data before publication.[10]

[9] Actually, the sum would be somewhat larger than the provincial sums because energy data from Tibet are not included in the data reporting.

[10] Involvement of outside organizations would of course require the execution of non-disclosure agreements to protect confidential information.