Working with Chinese Government Data Sets: Potential Issues and Solutions

Edward Junhao Lim
Library, New York University Shanghai, Shanghai, China.
E-mail address: edward.lim@nyu.edu

Jennifer Stubbs
Library, New York University Shanghai, Shanghai, China.
E-mail address: jas58@nyu.edu

Qinghua Xu
Library, New York University Shanghai, Shanghai, China.
E-mail address: qinghua.xu@nyu.edu

Abstract:

This paper will provide an overview of selected official statistics and data sets collected in the People's Republic of China, with a particular focus on economic and financial data on mainland China. It will explore options beyond the official statistics, such as CNKI, China Data Online, and CEIC CDMNext, and discuss the strengths and weaknesses of these databases.

It will also review the existing arguments surrounding the use of official Chinese statistics and how researchers need to identify any issues behind their data choices. The key takeaway is to avoid generalizing and judging the quality of Chinese data based on exceptions, but instead to approach official Chinese data with a nuanced approach.

Keywords: Government data; China; Statistical yearbooks; Economic statistics; National Bureau of Statistics.

Introduction

Today, China saturates the news cycle because of its recognized status as an economic superpower. China's official statistics are crucial for the world economy. Recently, there have been fears about China's economic impact on global markets. In comparison to the scrutiny of the U.S. economy, which quarterly, monthly, and even weekly indicators are continuously tracked, there is a lack of attention and knowledge on indicators of the Chinese economy.

This paper provides an overview of selected official statistical information collected in the People's Republic of China, with a particular focus on economic and financial data on mainland China[1]. It discusses key data sources, and comments on the limitations of some of the sources. Next is a short review of the arguments surrounding the use of official Chinese statistics. The goal is to illuminate the "black box" of official Chinese data and its sources.

National Bureau of Statistics of China

The *China Statistical Yearbook* is published by the China Statistics Press annually. The National Bureau of Statistics (NBS; 国家统计局[2]) compiles it. The *China Statistical Yearbook* started in 1981 and the bilingual version appeared with the 1994 edition. Later, separate English editions from 1981 to 1993 were published.

The year indicated on the title refers to the year it is published, usually in the third quarter (around September or October). The data covered is for the year before[3]. For example, the data found in the *China Statistical Yearbook 2018* is for 2017.

The statistical yearbooks are freely available online, as image files. This might work if one requires a glance, but it is not useful for a researcher who needs to manipulate the data on a spreadsheet. The print book contains a CD-ROM of both image files (in .PNG), and Excel spreadsheet files (in .XLS).

For a comprehensive listing of statistical publications at all levels of government, use the catalog by China Statistics Press (National Bureau of Statistics of China, n.d.-b).

The hierarchical nature of statistics in China
The Statistical Law of China requires local governments above the county level to publish statistical yearbooks, but they are usually in Chinese only – except for specific cities such as Beijing, Shanghai, Guangdong, Chongqing and major provinces who have an English edition as well. Provincial-level statistical publications will provide more detail than a national-level publication (Xue, 2004, p. 10).

Ministry-level and local level statistics
Parallel to the national level, there are also ministry-level statistical surveys, as well as local statistical surveys conducted by local governments (Xue, 2004). At the ministry-level, it is usually titled as a 'yearbook'; the term "statistical yearbook" is usually reserved for NBS publications. To quickly find the statistics section of the ministries or provincial-level statistics departments, the Chinese version of the NBS website is the right starting place.

---

[1] This short conference paper is unable to explain *all* Chinese government datasets; the Special Administrative Regions of Hong Kong and Macau, as well as Taiwan (also known as the Republic of China), is excluded from this paper.

[2] While we have included the Chinese and English titles or names for many of the sources mentioned in this paper, for accurate identification, the Chinese version should be used. We have noted several instances where the English title or name may have changed over time. We are following the style in China, where the Chinese version of any document has the force of law.

[3] The only exception is for its first edition in 1981, with the data from 1981.

Industry, sector, and topical statistics
In addition to the types of publications, some focus on a particular industry, sector, or topic across time. Holz (2013) notes that these are published by government ministries or commercial associations in that sector. Holz (2013, pp. 15–19) has compiled a thorough list of statistical yearbooks on specialized topics, published by NBS. It contains detailed explanations such as publication month, content coverage, and date ranges. An industry example is the *Almanac of China's Finance and Banking* (中国金融年鉴) – compiled by the China Finance Society, and published by China Finance Publishing House (中国金融年鉴杂志社有限公司). It is available in separate language editions. On the subject of women in China, there is the *Statistics on Chinese Women* (中国妇女统计资料) (1949–1989) covering a broad range of issues such as fertility, school enrolment, and awardees in science and technology fields.

Time-series data (e.g., 20 years of the same data over time)
"General yearbooks" are also published by the same bodies. These tend to be in Chinese only. Occasionally titles and section headers are translated in English, but the main textual content remains in Chinese. Note that these publications tend to be more textual with supporting statistics. An example is the *China Compendium of Statistics* (新中国六十年统计资料汇编) (1949–2008). Its national perspective includes the two Special Administrative Regions (Hong Kong and Macau).

It must be noted that data availability before 1978 can be problematic since this is known as the pre-reform period. Also, the quality of the data is likely to be poor as data for specific years (i.e., 1952–1977) were put together retrospectively (Holz, 2013).

Browsing experience: Native English version versus a dynamically translated English version
The English version of the NBS website is not as current as the Chinese version. For example, news announcements are several years out of date and the copyright year stated at the footer is 2012. Entering the English version feels like traveling back to a particular moment in time. With the convenience of browser-based translators (i.e., Google Translate in Chrome, Firefox), one would think that it is possible to navigate the Chinese version with little difficulty. Unfortunately, a researcher with no knowledge of Chinese could grow frustrated quickly with a dynamically translated NBS webpage. The translated text overflows in the navigation menus, making reading challenging and relevant sections obscure. In instances of legible translations, the idiomatic meaning is often lost.

Searching across yearbooks and indicators
NBS also has a "National Data" web portal (国家数据) available in both languages (National Bureau of Statistics of China, 2019). If following a publication-level citation, a researcher will require time to discover and explore what indicators are relevant for their research. One may wish to search for specific indicators. Most of the commercially available options listed below have this feature, as controlled vocabulary is not used consistently.

Take note that when searching in English, this increases the number of irrelevant results, or may result in no results because the search term is not found in the translated version – an extremely frustrating experience since the researcher is not clear if the indicator does not exist, or if it is a non-preferred variant.

There is a higher amount of similarity between the different languages than their organization website, but minor differences remain. For example, the English version of the Publications webpage in the National Data web portal offers (i) the *China Statistical Yearbook*, (ii) the *Statistical Communiqué*, and (c) *BRICS Joint Statistical Publication* (National Bureau of Statistics of China, n.d.-a). However, there is an additional publication on the corresponding Chinese version page (出版物): the *International Statistical Yearbook* (国际统计年鉴). This book is a compilation of data from World Bank, International Financial Statistics from IMF (International Monetary Fund) Data, and United Nations.

Key Data Sources

The following sources require a subscription or fee. They have mostly repackaged data from the statistical yearbooks, with varying coverage in content and turnaround time for updates.

If government data is freely available, why do institutions need to subscribe to commercial platforms? Because there are usually enhancements: indexing, translation, machine-encoded text (OCR), and normalized data (which allows for comparisons between variables and geographies). Bear in mind that there is a time lag for data updates.

CNKI (China National Knowledge Infrastructure; 中国知网)
CNKI has a subscription-based portal called *China Data Insights* (cdi.cnki.net), including statistical yearbooks. The full list of 1,086 'titles' is an excellent strategy if you have a specific title in mind (say, from a citation you have obtained from another source). Alternatively, you can browse by one of 18 topics. More importantly, it allows the user to search for specific entries ('tables') and figures ('indicators').

China Data Online (中国数据在线)
This is a similar subscription-based portal commonly known as *China Data Online* (CDO; also called *All China Data Center*), owned by All China Marketing Research (ACMR). Before October 2018, they were affiliated with the University of Michigan and Inter-university Consortium for Political and Social Research (ICPSR) (Leung, 2018).

Its greatest strength is their claim to have authorization and a direct relationship with NBS. ACMR claims to be a spinoff of NBS, and thus the data partner to international users. Also, it claims to have a complete collection for the provincial level statistical yearbooks, in both English and Chinese language, in a digital format (China Data Center, 2018).

CDO suffers from an outdated interface. Its China Geo-Explorer (CGE) relied on Flash technology until a recent upgrade in March 2019 to support HTML5.

CEIC & CDMNext
The third option is CEIC, owned by a private company called CEIC Holdings Limited. They were part of Euromoney Institutional Investor Group. In 2018, Global Market Intelligence Division (GMID) was acquired jointly by Caixin Global and Citic Capital. (The other unit within GMID is *EMIS*.) As of 2019, Caixin Global news articles are citing CEIC data, while CEIC now lists the "Caixin China New Economy Index" (CEIC, 2018).

CEIC's platform is known as *CDMNext*. It contains *China Premium Database*. Its coverage begins with 1949. It does not explicitly name its sources anywhere on its website or PDF brochures; our understanding from the support team is that their primary sources include

China's ministries and other industry bodies, such as the China Association of Automobile Manufacturers (CAAM). In comparison with the databases as mentioned earlier, CEIC updates its datasets more frequently.

Peking University (PKU) China Survey Data Archive (CSDA;北京大学中国调查数据资料库)

PKU's data repository – Open Research Data (北京大学开放研究数据平台) – is noteworthy because the various datasets it collects are frequently cited and mentioned in publications (Peking University Library, 2015). Some of the important longitudinal studies include:

- *China Family Tracking Survey* (CFPS; 中国家庭追踪调查): This is a national survey at the family unit level. The CFPS baseline survey sample covers 649 villages in 161 districts and counties in 25 provinces and cities nationwide, including 15,000 household samples and all family members in the sample households. There has been an attempt to gather follow-up visits every two years.

- *China Health and Pension Tracking Survey* (CHARLS; 中国健康与养老追踪调查): This is a national level family and individual survey aimed at those aged 45 and over. Launched in 2011, it covers 150 districts and counties, 450 villages, and about 17,000 members of the approximately 10,000 households, with follow-up visits every two years.

- *China's Old Age Health Influencing Factors Tracking Survey* (CLHLS; 中国老年健康影响因素跟踪调查): This is a national level survey aimed at those aged 65 and over, and includes their 35–64-year-old adult children.

Access to the datasets would require signing up for an account in order to request permission. Besides contact details, it requires an organizational affiliation, organizational role, research areas, highest educational qualifications, and job title. There is usually no fee involved in the process.

Other Noteworthy Options

The landscape when it comes to Chinese datasets can be chaotic at times. Here are some suggestions that can be a good starting point for researchers:

- Wang, Chen, & Richards (2018, p. 652) has a list of 7 provincial-level open government data portals, including Hong Kong and Macau. They note the lack of a national level open government data portal in China; unlike data.gov by the U.S. government and data.gov.uk by the U.K. government. However, 19 Chinese local level portals have been built, as a result of policies promoting the development of big data at the highest level in China's government.

- At the point of writing, the China Open Data Index (中国开放数林指数), a project by the DMG Lab of Fudan University, lists over 70 Chinese local open government data portals on their website (DMG Lab Fudan University, 2017).

- The Organisation for Economic Cooperation and Development (OECD) does a "periodic review of the Chinese economy" and is published as *OECD Economic Surveys* (2019). The book is freely available to read online and contains a good number of graphs and figures that can be viewed in XLS or CSV format.

- New York University Shanghai has a *Chinese Datasets Archive* portal, which "aims to serve as a starting point for students and scholars to search for data on China" (NYU Shanghai, 2018).

Potential Issues and Challenges

Bill Gross, a bond-fund manager, once called China "the mystery meat of emerging-market countries" (Pi, 2017). It is no exaggeration that official Chinese statistics have a reputation issue, going by the many adverse news reports (in the Western media) about the unreliability and inaccuracy of the country's official statistics.

Cooking the Books
Plekhanov (2017) reviewed the English-language scholarly research on Chinese official statistics from the early 2000s and identified three main problems that researchers have reported. In decreasing order of prevalence: (a) inadequate methodology, (b) underreporting, and (c) data manipulation. For example, China's National Health Commission attribute flu deaths different compared to other countries because its methods do not follow the practice used elsewhere. Thus, the "unusually low" flu deaths can create a misunderstanding about how Chinese people do not die because of flu, which in turn affects people's behavior (Teng, 2019).

Lack of Transparency and Independence
Researchers are also paying attention to the lack of transparency offered by NBS. This is a structural issue since the heads of NBS and agencies are political appointees. (This is not to say that they are politicians; many are professional statisticians.) Their taskmasters are ultimately politicians since the NBS is a government department under the direct control of the Party (Bernanke & Olson, 2016). The current director of NBS, Ning Jizhe (宁吉喆), is also Vice Chairman of the National Development and Reform Commission (Bloomberg News, 2016).

Other criticisms include data series inconsistencies – because of changes in coverage and re-definition – and data divergence. This is when similar statistical indicators do not correspond with each other. An often-cited example is that China does not publish GDP price deflators. This results in problems of data inconsistency since researchers need to derive implicit GDP deflators from the nominal GDP figures (Orlik, 2012).

As a caveat, many of the studies covered by Plekhanov (2017) focus on just three economic indicators, which are GDP, unemployment, and industrial production. Data on the unemployment rate appeared to be the most problematic of the three. We must be careful not to paint *all* Chinese official statistics as unreliable and inaccurate based on past cases.

Debunking the Falsification Theory
In defense of NBS, Orlick (2012) has advocated for sensible use of Chinese data. He uses Holz's refutation of systematic manipulation to counter Rawski, Wu, and Maddison's

contentions against China's reported GDP in 1998. Orlik contends China continues to catch up with the rapid growth of new sectors in China's economy, which requires innovation by NBS in their survey and reporting tools. These include sample calibration, directly collecting data, and adopting new or international techniques, such as IMF reporting standards.

For example, Fernald, Malkin, & Spiegel (2013) – on behalf of the Federal Reserve Bank of San Francisco – tested the accuracy of China's GDP data by comparing them with alternative indicators[4] "over a long time span that seemed less subject to manipulation." They concluded that the data were consistent and there was no deviation – at least from 2000 to 2012. This conclusion applied even when using data not produced by the Chinese authorities. However, they did not rule out the possibility that the Chinese GDP could have been over reported over time.

Orlik (2012) admonishes the Ministry of Human Resources and Social Security's Unemployment quarterly report's limited population, preferring their Public Employment Services Labor Market Supply and Demand Situation quarterly report (since 2001) for capturing migratory labor, job openings, and motive for job changes, across 100 cities.

Finally, Orlik (2012) critiques the secrecy of most offices for methods and weights (such as the Consumer Price Index or GDP deflator), leading citizens to distrust and underreport. Wildau (2015) has captured part of the contention among economists: Carsten Holz is seen as the defender of China's official data and claims that China's nominal GDP figures are mostly accurate. At the opposing end, Harry Wu remains skeptical of the Chinese industrial GDP data. However, both agree that data collection by Chinese officials has to improve – be it the publication of the GDP deflator, or minimizing political interferences.

Translation Quirks
The same can be said for the websites of many government bodies. An unfortunate example is the English version of the Ministry of Water Resources of the PRC (中国人民共和国水利部) website (n.d.). It uses images to display words instead of text, thus defeating browser-translation. Sections of the website still require Adobe Flash, which is no longer supported by most browsers. Unfortunately, the browsing experience is inferior to the Chinese version.

One case in point of the discrepancy between the two versions is statistical data on the People's Bank of China (中国人民银行) website. The historical data coverage goes back to 1999 in the Chinese version, but only 2006 in the English version (The People's Bank of China, 2019b, 2019a). (At the point of writing, the 2019 data is available on the Chinese version, but not yet available on the English version.) Upon closer examination, one would notice that links on both versions navigate to the same landing page.

Conclusion

As China continues to improve its business environment, the pressure for government information disclosure will persist (Caixin Global, 2019). Similar to Holz, we are optimistic about Chinese data because of the "abundance of statistics published by China's statistical authority and Chinese government departments" (Holz, 2013, p. 25). The caveat is that the researcher has to understand the variety of sources and "what the data mean."

---

[4] See Appendix B for an elaboration of alternative indicators used by other researchers to verify official Chinese statistics.

Koch-Weser's advice to U.S. policymakers is relevant to when we provide advisory and consultation to our researchers: Chinese statistics cannot be seen to be uniformly reliable or unreliable. Rather users should adopt a nuanced and contextual approach dependent on the time and unit of measurement (2013) and the agency reporting the data. Most importantly, researchers should be able to identify the reason behind any data issues. Librarians and information professionals have a role to play in providing a menu of options concerning China's economic and financial records. We also have a responsibility to guide and educate researchers who desire to find and make sense of data with Chinese characteristics.

Acknowledgments

References

Bernanke, B. S., & Olson, P. (2016, March 8). China's transparency challenges. Retrieved April 29, 2019, from Brookings website: https://www.brookings.edu/blog/ben-bernanke/2016/03/08/chinas-transparency-challenges

Bloomberg News. (2016, February 27). China Names Close Adviser to Premier Li as New Statistics Chief. Retrieved May 23, 2019, from https://www.bloomberg.com/news/articles/2016-02-27/china-names-close-adviser-to-premier-li-as-new-statistics-chief

Caixin Global. (2019, April 23). Editorial: To Improve China's Business Environment, Access to Government Information Must Be Widened. Retrieved April 26, 2019, from https://www.caixinglobal.com/2019-04-23/editorial-to-improve-chinas-business-environment-access-to-government-information-must-be-widened-101407650.html

CEIC. (2018, June 21). Caixin China New Economy Index added to CEIC. Retrieved May 6, 2019, from CEICdata website: https://www.ceicdata.com/en/blog/caixin-new-economy-index

Chen, W., Chen, X., Hsieh, C.-T., & Song, Z. (Michael). (2019). A forensic examination of China's national accounts. *Brookings*. Retrieved from https://www.brookings.edu/bpea-articles/a-forensic-examination-of-chinas-national-accounts

China Data Center. (2018, October 1). A comparison between ACMR and other data sources. Retrieved May 6, 2019, from China Data Institute website: https://chinadatacenter.net/Announcement/AnnouncementContent.aspx?id=2496

Chorzempa, M. (2016, June 27). P2P Series Part 1: Peering Into China's Growing Peer-to-Peer Lending Market. Retrieved May 20, 2019, from PIIE website: https://piie.com/blogs/china-economic-watch/p2p-series-part-1-peering-chinas-growing-peer-peer-lending-market

DMG Lab Fudan University. (2017). China open data index. Retrieved May 19, 2019, from http://ifopendata.fudan.edu.cn/method

Fernald, J., Malkin, I., & Spiegel, M. (2013, March 25). On the Reliability of Chinese Output Figures. Retrieved April 30, 2019, from https://www.frbsf.org/economic-research/publications/economic-letter/2013/march/reliability-chinese-output-figures

Holz, C. A. (2013). Chinese statistics: classification systems and data sources. *Eurasian Geography and Economics*, *54*(5–6), 532–571. https://doi.org/10.1080/15387216.2014.908312

Koch-Weser, I. N. (2013). *The Reliability of China's Economic Data: An Analysis of National Output* (p. 44). Retrieved from https://www.uscc.gov/sites/default/files/Research/TheReliabilityofChina'sEconomicData.pdf

Leung, R. (2018, October 3). With little explanation, "U" abruptly cuts ties with China Data Center | The Michigan Daily. Retrieved May 23, 2019, from The Michigan Daily website: https://www.michigandaily.com/section/research/little-explanation-u-m-abruptly-cuts-ties-china-data-center

McMahon, J. (2017, November 17). How An App Is Stopping Pollution In China. Retrieved May 23, 2019, from Forbes website: https://www.forbes.com/sites/jeffmcmahon/2017/11/17/how-an-app-is-stopping-pollution-in-china

Ministry of Water Resources. (n.d.). Ministry of Water Resources the People's Republic of China. Retrieved May 23, 2019, from http://www.mwr.gov.cn/english

Minter, A. (2014, March 6). China's Li Doesn't Believe His Own Numbers. Retrieved May 22, 2019, from https://www.bloomberg.com/opinion/articles/2014-03-05/china-s-li-doesn-t-believe-his-own-numbers

National Bureau of Statistics of China. (2019). National Data. Retrieved May 27, 2019, from http://data.stats.gov.cn/english

National Bureau of Statistics of China. (n.d.-a). Publications. Retrieved May 27, 2019, from http://data.stats.gov.cn/english/publish.htm?sort=1

National Bureau of Statistics of China. (n.d.-b). 国家统计局>>统计出版 Guójiā tǒngjì jú >>tǒngjì chūbǎn [National Bureau of Statistics>>Statistical Publishing]. Retrieved May 23, 2019, from http://www.stats.gov.cn/tjzs/tjsj/tjcb

NYU Shanghai. (2018). Datasets | Data Science. Retrieved May 14, 2019, from https://datascience.shanghai.nyu.edu/datasets

OECD. (2019). *OECD Economic Surveys: China 2019*. https://doi.org/10.1787/eco_surveys-chn-2019-en

Orlik, T., 1978-. (2012). *Understanding China's economic indicators : translating the data into investment opportunities*. Upper Saddle River, N.J.: FT Press.

Owyang, M. T., & Shell, H. G. (2017). China's Economic Data: An Accurate Reflection, or Just Smoke and Mirrors? Retrieved April 29, 2019, from https://www.stlouisfed.org/publications/regional-economist/second-quarter-2017/chinas-economic-data-an-accurate-reflection-or-just-smoke-and-mirrors

Peking University Library. (2015, 2017). Peking University Open Research Data Platform. Retrieved May 23, 2019, from http://opendata.pku.edu.cn

Pi, X. (2017, March 22). Quicktake: China's Economic Data. Retrieved April 29, 2019, from Quicktake: China's Economic Data website: https://www.bloomberg.com/quicktake/chinas-economic-data

Plekhanov, D. (2017). Quality of China's Official Statistics: A Brief Review of Academic Perspectives. *Copenhagen Journal of Asian Studies*, *35*(1), 76–101. https://doi.org/10.22439/cjas.v35i1.5400

Teng, J. X. (2019, February 21). Why Aren't People in China Dying of the Flu? Retrieved May 18, 2019, from https://www.caixinglobal.com/2019-02-21/why-arent-people-in-china-dying-of-the-flu-101382286.html

The Economist. (2010, December 11). Asia: Keqiang ker-ching; China's economy. *The Economist; London*, *397*(8712), 54.

The Institute of Public & Environmental Affairs (IPE). (2010). Home. Retrieved May 23, 2019, from http://wwwen.ipe.org.cn

The People's Bank of China. (2019a). Home-Survey & Stat. Retrieved May 23, 2019, from

http://www.pbc.gov.cn/english/130730/index.html

The People's Bank of China. (2019b). 中国人民银行-调查统计司-统计数据 Zhōngguó rénmíng yínháng-diàochá tǒngjì sī-tǒngjì shùjù [People's Bank of China - Survey and Statistics Division - Statistics]. Retrieved May 19, 2019, from http://www.pbc.gov.cn/diaochatongjisi/116219/116319/index.html

Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly*, *35*(4), 644–656. https://doi.org/10.1016/j.giq.2018.10.006

Wildau, G. (2015, September 28). China data: Making the numbers add up. Retrieved March 11, 2019, from Financial Times website: https://www.ft.com/content/cb446e10-6057-11e5-97e9-7f0bf5e7177b

Wildau, G., Cadman, E., Yang, Y., & Kavanagh, L. (2019). China's economy: GDP growth, interest rates and inflation statistics. Retrieved May 27, 2019, from https://ig.ft.com/sites/numbers/economies/china

Xinhua. (2007, December 13). Chinese NGO blacklists 4,000 companies for bad air pollution record - People's Daily Online. Retrieved May 23, 2019, from http://en.people.cn/90001/90776/6320616.html

Xu, X. (2013). Government publications in Chinese academic libraries. In *Academic libraries in the US and China: Comparative studies of instruction, government documents and outreach* (pp. 131–164). https://doi.org/10.1016/B978-1-84334-691-3.50004-4

Xue, S. (2004). China's statistical system and resources. *Journal of Government Information*, *30*(1), 87–109. https://doi.org/10.1016/j.jgi.2004.07.002

Ye, Y. (2014). *The ALA guide to researching modern China*. Chicago: ALA Editions, an imprint of the American Library Association.

Appendix A

Government Documents

This is not the focus of this paper, but we feel it is essential to mention. The *ALA Guide to Researching Modern China* lists out five significant types of government documents published at all levels of the Chinese government, in decreasing level of importance: (i) National People's Conference documents, (ii) speeches of national leaders, (iii) government/party notices and decisions, (iv) internal documents at government agencies and branches, (v) government annual reports or white papers (Ye, 2014, pp. 84–85). Xu (2013) explained that that various government publications are called "red header documents," and they usually originate from governmental organizations like the State Council and its affiliated institutions, and local governments.

Lawinfochina is a subscription platform to access most of these government documents. Hosted by Chinalawinfo Co. Ltd. (北京北大英华科技有限公司), PKU established it in association with the university's Legal Information Center. It is also known as their "English Translation Database," or more recently, *pkulaw*, It contains not only the full text of government documents going back to 1949 but also a section-by-section translation.

Alternatively, CNKI has the China Government Gazettes Full-text Database (中国政报公报期刊文献总库). The platform interface is currently available in Chinese.

The *Chinese Government Public Information Online* (中国政府公开信息整合服务平台) is noteworthy because it is an attempt to provide various types of information from the municipal, provincial, or autonomous regional government organizations (Xu, 2013). It provides a static screenshot of those government organization websites, a text-only pop-up, as well as a direct link to the webpage. The platform allows one to filter to "statistics" (统计信息); however, many are declarative headlines from the local governments, so do not expect tabular data. At the time of writing, it was only available in Chinese.

Appendix B

Alternative Indicators

Researchers should consider using alternative or proxy indicators if they are unable to verify the accuracy of the statistics at hand.

For a short list of such alternative indicators, Koch-Weser (2013, p. 27) compiled a table of indicators listing category, the unit of measurement, and source. For example, the purchasing managers index (PMI, now a product of Markit; see "Purchasing Managers' Index®") collected by HSBC is seen as an independent effort to complement the work of Chinese statisticians.

Li Keqiang Index
Li Keqiang – currently China's Premier but when he was party secretary of Liaoning province – was quoted as using three metrics to judge economic performance in the country: (a) electricity (or energy) consumption, (b) rail freight volume (or railway cargo volume), and (c) new bank loans (Minter, 2014). Orlik (2012, p. 10) has explained that this was revealed during the Wikileaks scandal in 2010. *The Economist* (2010) popularized this by calling it the "Li Keqiang Index," and this has continued to be used in media, including *Financial Times* (Wildau, Cadman, Yang, & Kavanagh, 2019).

Luminosity as a Proxy for Economic Activity
Chen, Chen, Hsieh & Song (2019) have also used nighttime lights data from U.S. National Oceanic and Atmospheric Administration (NOAA) as an estimation of economic activity in China. It is claimed that using satellite data to measure luminosity – the intensity of human-made night lights – is a reliable method and proxy to measure economic activity (Owyang & Shell, 2017).

However, we need to temper our researchers' expectations when it comes to alternative sources of data; many are still limited in terms of coverage and scope. Even with alternative indicators, they can be subject to criticism, and may not be any closer to the 'actual' reality if compared with the official indicators.

New Business, Financial, and Environmental Indicators Needed
There are sectors of China's 'new' economy that are not adequately captured. For example, understanding the Chinese consumer market requires more data to be collected about service-oriented businesses like restaurants and karaoke bars (Pi, 2017). Thankfully, there is a growing attempt to capture the online market with China's reputation as an online-first marketplace. For example, WDZJ.com (网贷之家; or the "Home of Online Lending") is a private company in Shanghai, collecting data about the peer-to-peer lending industry in China. Their data matched China's central bank's figures for 2015 (Chorzempa, 2016).

Another example is the Institute of Public and Environmental Affairs (IPE), which collects government and corporate environment information in order to build their environment database (2010). They are a non-profit organization operating from Beijing. Established in 2006, they are now seen as the authority when it comes to environmental governance in China (McMahon, 2017; Xinhua, 2007).