NBER WORKING PAPER SERIES

CHINA'S GDP GROWTH MAY BE UNDERSTATED

Hunter Clark
Maxim Pinkovskiy
Xavier Sala-i-Martin

Working Paper 23323
http://www.nber.org/papers/w23323

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2017

We are very grateful to Michelle Jiang, Lauren Price and Rachel Schuh for superb research assistance. This paper reflects solely the opinions of the authors and not necessarily of the Federal Reserve Bank of New York, the Federal Reserve System, or the National Bureau of Economic Research.

China's GDP Growth May be Understated
Hunter Clark, Maxim Pinkovskiy, and Xavier Sala-i-Martin
NBER Working Paper No. 23323
April 2017
JEL No. F0

ABSTRACT

Concerns about the quality of China's official GDP statistics have been a perennial question in understanding its economic dynamics. We use data on satellite-recorded nighttime lights as an independent benchmark for comparing various published indicators of the state of the Chinese economy. Using the methodology of Pinkovskiy and Sala-i-Martin (2016a and b), we exploit nighttime lights to compute the optimal weights for various Chinese economic indicators in a best unbiased predictor of Chinese growth rates. Our computations of Chinese growth based on optimal weightings of various combinations of economic indicators provide evidence against the hypothesis that the Chinese economy contracted precipitously in late 2015, and are consistent with the rate of Chinese growth being higher than is reported in the official statistics.

Hunter Clark
Federal Reserve Bank of New York
33 Liberty St, New
York, NY 10045, USA
Hunter.Clark@ny.frb.org

Maxim Pinkovskiy
Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045
maxim.pinkovskiy@ny.frb.org

Xavier Sala-i-Martin
Department of Economics
Columbia University
420 West 118th Street, 1005
New York, NY 10027
and NBER
xs23@columbia.edu

1 Introduction

We are interested in understanding recent growth rates in China and whether they coincide with the officially published statistics. There is a large literature doubting the accuracy and veracity of officially published Chinese GDP data, often claiming that official growth rate considerably overstates actual growth. In a seminal article, Rawski (2001) argues that the Chinese economy might have grown at 2% or less per year during 1997-2001 instead of the 7.1% as officially claimed. Other researchers, in particular Adams and Chen (1996), Maddison and Wu (2007) have shared this skepticism by comparing official Chinese growth estimates to growth rates of various inputs into production, such as energy, steel and cement. On the other hand, Holz (2013) and Perkins and Rawski (2008)[1] have claimed that the Chinese data are generally accurate. Recently, Chinese statistics have been at the source of additional controversy following Wikileaks' publication of the premier, Li Keqiang (then still a provincial governor), admitting to an American diplomat that he monitored provincial economic activity by the simple arithmetic average of the growth rates of electricity production, railroad freight and bank loans, and that the official statistics were "man-made" and "for reference only." *The Economist* has been reporting the Li Keqiang index to this day. However, all of the above proposals for measuring the growth rates of the Chinese economy depend on assumptions about the relationships between various macroeconomic proxies and economic activity, many of which are difficult to evaluate.

In this paper, we attempt to transcend the problem of understanding which theory is right by using a variable that can act as an "impartial referee" for the macroeconomic proxies. Pinkovskiy and Sala-i-Martin (2016a and b) show that if we can find a variable whose measurement errror is independent of the measurement errors of the macroeconomic variables that forecasters typically aggregate to predict economic activity, we can a) test the quality of these variables, and b) obtain the optimal weights on these variables in a best unbiased linear predictor of unobserved true income. We argue that such a variable is satellite-recorded nighttime lights (Elvidge et al. 1993, Henderson et al. 2012, Pinkovskiy and Sala-i-Martin 2016a and b). We believe that this is a reasonable assumption because errors in nighttime lights come from variation in weather patterns as well as from variation in satellite quality over time. On the other hand, errors in GDP and other official series come from misreporting by individuals, firms and other institutions, as well as from methodological choices of statistical offices. Therefore, there is little reason to believe that these errors have anything to do with each other.

Under our crucial assumption and using data across Chinese provinces and over time for the period 2004-2013, a regression of log nighttime lights on the macro proxies as well as on province and year fixed effects yields coeficients on the proxies that are proportional to the proxies' optimal weights (see the mathematical

[1] This is the same Rawski as Rawski (2001).

framework in Section 2.1). We find that the Li Keqiang variables are significant predictors of nighttime lights growth, but that they receive radically different weights. In particular, bank loans receive considerably more that one-half of the total weight, while railroad freight receives much less than one-third. Electricity production also typically receives a large and statistically signficant coefficient, but it may be biased upward if our core identification assumption fails. Formally testing that the coefficients on the three Li Keqiang variables are equal rejects this null hypothesis in most specifications, and the failures to reject that we find are driven by large standard errors rather than similar coefficients. Since the growth rate of loans is larger and more stable over time than the growth rates of electricity and (especially) freight, this finding is important for predicting true unobserved Chinese GDP growth. Moreover, we find that adding log GDP to the the separate components of the Li Keqiang index does not materially change the coefficients on the Li Keqiang variables, and the coefficient on log GDP is only marginally significant. We also find that conservative adjustments for the potential correlation between the measurement error in nighttime lights and the measurement error in electricity do not change our conclusions concerning the importance of loans in explaining nighttime lights. We perform similar regressions including other variables besides the Li Keqiang components that have been used by analysts to predict Chinese growth (e.g. retail sales, passenger traffic and floor space under construction, among others) and find that bank loans continue to be important in explaining nighttime lights, and therefore, in predicting Chinese economic growth.

For each regression that we run, we construct forecasts for the path of Chinese growth rates by taking weighted averages of the national growth rates of the macroeconomic proxies as implied by the regression coefficients. A remaining problem is that the regression coefficients are proportional, but not identical, to the variables' optimal weights, and any intercept is unidentified. Hence, we normalize the resulting growth paths by regressing the official GDP growth path on the weighted average for the 2004-2012 period and using the predicted values of this second regression (in and out of sample) as our predicted GDP growth series. We obtain standard errors for the predicted paths of Chinese GDP growth by parametric bootstrapping from the asymptotic distribution of the estimated coefficients from the provincial-level regression. We find that GDP growth in 2015 Q4, a time when the financial press was awash with stories about a "hard landing" of the Chinese economy, is somewhat higher, but quite close to the officially reported rate, with a 95% confidence interval that precludes growth rates consistent with a sharp slowdown of the Chinese economy at this time. We also obtain this finding in our most comprehensive specifications that include different sets of macro proxies corresponding to the various indices used by market participants to predict the Chinese economy. Our finding is intuitive because the nighttime lights suggest a particularly high weight for bank loans in the best unbiased linear predictor of Chinese economic growth, and, unlike freight and electricity growth, which fell during 2015, bank loan growth remained high over the course of that time.

The rest of the paper is organized as follows. Section 2 describes the mathematical framework for using nighttime lights to construct optimal indices of predictor variables. Section 3 presents a description of the data. Section 4 describes our methodology for converting the index values into forecasts. Section 5 presents the baseline results. Section 6 presents results with covariates from other indices of the Chinese economy. Section 7 concludes.

2 Mathematical Framework

2.1 Outline of the Main Approach

This subsection summarizes the mathematical framework of Pinkovskiy and Sala-i-Martin (2016a and b). Formally, we assume the model

$$\begin{aligned} y_i^L &= f_i\left(y_i^*\right) + \varepsilon_i^L \\ y_i^G &= \alpha^G + \beta^G y_i^* + \varepsilon_i^G \\ y_i^E &= \alpha^E + \beta^E y_i^* + \varepsilon_i^E \end{aligned}$$

with

$$cov\left(y_i^*, f_i\left(y_i^*\right)\right) \neq 0 \tag{A0}$$

$$E\left(\varepsilon_i^G | y_i^*\right) = E\left(\varepsilon_i^E | y_i^*\right) = E\left(\varepsilon_i^L | y_i^*\right) = 0 \tag{A1}$$

$$E\left(\varepsilon_i^L \varepsilon_i^G | y_i^*\right) = E\left(\varepsilon_i^L \varepsilon_i^E | y_i^*\right) = 0 \tag{A2}$$

where y_i^L denotes nighttime lights for location i, y_i^* denotes unobserved true income for location i, and y_i^G and y_i^E are the candidate measures. Note that the only restriction on f_i is that it satisfies assumption $A0$. Let us consider an estimator of true income

$$z_i = \eta + \gamma^G y_i^G + \gamma^E y_i^E$$

and let γ_G^* and γ_B^* be the MMSE-minimizing weights for this this estimator. Then, if we run the simple linear regression

$$y_i^L = \alpha + b^G y_i^G + b^B y_i^E \tag{1}$$

we can recover

$$\frac{b^E}{b^G + b^E} = \frac{\gamma_E^*}{\gamma_G^* + \gamma_E^*}$$

The proof is provided in Pinkovskiy and Sala-i-Martin (2016a and b). These arguments are straightforwardly extended to more than two proxies, with the result for K proxies being that

$$\frac{b_i}{\sum_{j=1}^K b_j} = \frac{\gamma_i^*}{\sum_{j=1}^K \gamma_j^*}$$

We can straightforwardly extend this framework to assumptions A0-A2 holding conditional on covariates, such as location and time fixed effects.

2.2 Bounds on Potential Violations of Exogeneity of Errors in Lights

Our framework allows us to consider the possible effects of violations of our identification assumptions. In particular, one of the right hand-side variables of interest will be electricity consumption, which is likely jointly determined with nighttime lights. In this case, we may have

$$E\left(\varepsilon_i^L \varepsilon_i^E | y_i^*\right) > 0$$

The covariance of electricity and nighttime lights will then be given by

$$cov\left(y_i^L, y_i^E\right) = \beta^E cov\left(y_i^*, f_i\left(y_i^*\right)\right) + bias$$

where *bias* is greater than zero. An upper bound for the bias would entail

$$\begin{aligned} \beta^E &= 0 \\ bias &= cov\left(y_i^L, y_i^E\right) \end{aligned}$$

In this case, a straightforward correction for the bias will be to replace the empirical covariance of electricity and nighttime lights with zero in the regression (1). A less conservative correction may be to use other information to find a plausible lower bound on β^E. In particular, we observe that if assumption A2 held, and we had data on both electricity and bank loans $\left(y_i^{loan}\right)$, we would obtain

$$\frac{cov\left(y_i^L, y_i^E\right)}{cov\left(y_i^L, y_i^{loan}\right)} = \frac{\beta^E}{\beta^{loan}}$$

If we assumed that the value of β^E should be no less than the value of β^{loan}, then we could bound $cov\left(y_i^L, y_i^E\right)$ below by $cov\left(y_i^L, y_i^{loan}\right)$. This could be empirically implemented by running an IV regression of y_i^E on y_i^{loan}, instrumenting using y_i^L, and then computing

$$cov\left(y_i^L, y_i^E\right)_{true} = cov\left(y_i^L, y_i^E\right)_{estimated} / \hat{b}_{IV}$$

3 Data

3.1 Nighttime Lights Data

This section is very similar to analogous sections in Pinkovskiy and Sala-i-Martin (2016a and b).

Data on lights at night is collected by the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) satellite program and is maintained and processed by the National Oceanic and Atmospheric Administration (NGDC 2010). Satellites orbit the Earth, sending images of every location between 65 degrees south latitude and 65 degrees north latitude at a resolution of 30 arcseconds (approximately 1 square km at the equator) at 20:30 to 22:00 local time.[2] The images are processed to remove cloud cover, snow and ephemeral lights (such as forest fires) to produce the final product available for download at

http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html

The nighttime lights data is available from 1992 to 2013.

Each pixel (1 square kilometer) in the luminosity data is assigned a digital number (DN) representing its luminosity. The DNs are integers that range from 0 to 63. We construct our lights proxy for aggregate income by summing up all the digital numbers across pixels

$$\text{Lights}_{j,t} = \sum_{i=1}^{63} i * (\text{\# of pixels in province } j \text{ and year } t \text{ with DN} = i\ )$$

This formula has been used to aggregate the nighttime lights maps into lights-based indices for each country and year in nearly the entire literature on nighttime lights in economics, including Henderson,

[2] There are one or two satellites recording nighttime lights in each year, with an old satellite being retired and a new satellite being launched every few years. The satellites from which data is available are as follows: the satellite F-10 (in orbit 1992-1994), F-12 (1994-1999), F-14 (1997-2003), F-15 (2000-2007), F-16 (2004-2009), F-18 (2010-2013), and the VIIRS satellite (April 2012-present).

Storeygard and Weil (2012), Chen and Nordhaus (2011) and Michalopoulos and Papaioannou (2013, 2014). For years with multiple satellites available, we average the logarithms of our aggregate luminosity measure, following Henderson, Storeygard and Weil (2012).

It is very well established that lights are very strongly correlated with measures of economic activity, such as national accounts GDP, in levels and growth rates. Henderson, Storeygard and Weil (2012) provide these correlations, dramatic pictures of long-term differences in incomes (North vs. South Korea) as well as short-term fluctuations (the Asian financial crisis of 1997-8) reflected in lights. Michalopoulos and Papaioannou (2013, 2014) present evidence that nighttime light density in a sample of African villages is correlated with development indicators for these villages. Table II of our paper show that there are strong regression relationships between log nighttime lights per capita and log GDP per capita in Chinese provinces, conditional on province and year fixed effects.

Our paper is closest in spirit to Henderson, Storeygard and Weil (2012) and Chen and Nordhaus (2011) in that it also considers the problem of optimally combining measures of economic activity. However, instead of using nighttime lights as a component of such a measure, we use it as an auxiliary variable to help uncover the correlation structure between the measures we do wish to use in our index. We also consider a different type of predictor for true income that do either Henderson, Storeygard and Weil (2012) or Chen and Nordhaus (2011), which allows us to make fewer assumptions on the data generating processes that we consider.

There are also well-known problems with the relation between nighttime lights and economic development, which we need to take into account. Pixels with DN equal to 0 or 63 are top- or bottom-censored. The light data also are affected by overglow and blooming: light tends to travel to pixels outside of those in which it originates, and light tends to be magnified over certain terrain types such as water and snow cover (Doll 2008). Given that we will compute national-level estimates of aggregate lights, it is unlikely that these sources of error will be large enough or sufficiently correlated with important variables that they will confound our analysis. Another problem may be that satellites age in space and are eventually retired. Hence, they might give inconsistent readings from year to year, or new satellites may give fundamentally different readings from old ones. While some evidence of this problem exists, the mathematical framework in Section 2 suggests that our calculations are supported by assumptions that allow nighttime lights to have all of the data problems described above, so long as nighttime lights are correlated with true income. We will also address this problem by including year fixed effects (sometimes additionally interacted with cross-sectional variation) in all specifications.

3.2 Chinese Economic Data

We obtain data on a variety of macro proxies for the Chinese economy for each Chinese region (province, municipality or autonomous region) for the years 2004-2013 from the CEIC database for the years 2004-2013. We also have national series of all of these variables at a quarterly frequency from 2004 to 2015. The full list of these variables is as follows:

1. GDP
2. Electricity production
3. Railroad freight volume
4. Total value of loans made by banks
5. Total freight volume
6. Retail sales
7. Value added by industry
8. Floor space under construction
9. Real estate investment
10. Airport throughput
11. Exports
12. Imports
13. Total passenger traffic
14. Fixed asset investment
15. Steel production
16. CPI
17. Fixed asset investment price index
18. Retail price index
19. Export price index (only at national level)
20. Import price index (only at national level)

We merged this data at the region-year level with nighttime lights data from the DMSP-OLS satellite. Figure 1 presents a map of Chinese provinces for reference about locations, sizes and regions. It is well known that some Chinese regions are very small and urban (the independent municipalities of Beijing, Tianjin and Shanghai), and to this end, we combine Hebei, Beijing and Tianjin (one group) as well was Jiangsu and Shanghai (second group) and Sichuan and Chongqing (third group) to avoid any contamination from cities having a different lights to true income relationship than less urban areas do. In our baseline sample, there are 27 provincial aggregates. We also divide up the provincial aggregates into 3 regional groups, as is shown in the map. The groups correspond to a group of rich, coastal aggregates (including the cities of Beijing, Shanghai and Guangzhou), a group of inland but centrally located areas (Manchuria and some hinterland provinces), and a group of remote western provinces, such as Tibet, Xinjiang and Inner Mongolia, but also including some urban western provinces such as Sichuan (with Chongqing) and Shaanxi.

Figure 2 presents national growth rates of GDP, the Li Keqiang index, and its three components (electricity, freight and loans). We see, in particular, that freight traffic has experienced a sharp but temporary drop in 2008 and grew at a negative rate in 2013, while the other components of the Li Keqiang index have not experienced such a deceleration.

It is well-known in the China forecasting industry that the real GDP growth rates reported by the provinces are inconsistent with the official real GDP growth rates published by the central government. There are multiple hypotheses as to why this divergence occurs, although it is believed that provincial governors exaggerate growth figures, necessitating adjustment to the total. Figure 4 presents the time series of annual national growth rates for real and nominal GDP as officially reported by the government and as implied by a weighted average of the provincial reports. We observe that while there is the evident discrepancy between official and provincial real GDP growth rates, there is much less discrepancy (except in 2005 and 2007) in the nominal GDP growth rates, and the discrepancies in the nominal GDP growth rates are smaller than in the real GDP growth rates when they exist. We therefore find it reasonable to infer that, in computing national real GDP growth rates, the central government essentially takes the nominal GDP growth rates reported by the provinces as given, and then deflates them using a common deflator.[3] We therefore look at both nominal and real GDP in our analysis at the provincial level, although we only look at real GDP when making our forecasts at the national level.

Many financial industry participants outside China have been skeptical of Chinese official statistics, with many analysts believing that the official growth rate overstates true growth. Several private forecasters have created alternative indices that indeed suggest that official GDP statistics overstate growth. We take a

[3] In our estimation sample the sum of annual provincial nominal GDPs (in levels) average about 5.6% greater than annual national GDP, with a maximum of 6.7% in 2012 and a minimum of 3.5% in 2007.

convenience sample of such indices that have received coverage in the financial press (Orlik and Chen 2015). Figure 5 illustrate these indices' predictions. The Barclays index uses purchasing managers' indices (PMIs), which do not exist at the provincial level, and cannot be replicated in our analysis. The Bloomberg and the Capital Economics (CAP) indices use linear combinations of variables such as industrial production (value added of industry or VAI), freight, passenger traffic, electricity production, floor space completed and retail sales. The Lombard Street index recomputes the deflator for real GDP by including prices of fixed asset investment, imports, exports and consumer prices into the calculation. We see that all the four indices show shortfalls in growth relative to the official numbers. The Bloomberg index is roughly in agreement with the official statistics, while the Barclays and the CAP indices show growth modestly lower than the official statistics do (in the range of 4.8-5.5% relative to 6.9% for 2015 Q3), while the Lombard Street index is consistent with radically lower Chinese growth at 3% in the same time period. Additionally, we investigate the determinants in the index proposed by John Fernald et al. (2015), who look at extracting principal components from a broad set of potential determinants of growth.[4]

4 Forecast Methodology

Once we obtain optimal weights on various possible variables in the provincial-level regressions, we will need to turn them into forecasts of recent Chinese GDP growth. We do so by 1) computing the weighted average of the national growth series corresponding to the regressors in the provincial regressions, with the weights equal to the provincial regression coefficients, 2) regressing the official national quarterly real GDP series on this weighted average (which, recall, is also national and at quarterly frequency) during the period 2005-2012 (before the divergences between the forecasting indices and official GDP growth begins), and 3) using the fitted values of this regression as predictions of GDP growth rates in China both in and out of sample. We bootstrap this procedure 100 times by drawing index weight vectors from the asymptotic distribution of the provincial regression coefficients and computing a separate prediction with each vector. We present the prediction based on the point estimates of the provincial regression as our point estimate. Figures 10 and 11 show the predicted paths of Chinese growth using this forecast procedure, together with 95% confidence intervals and with the official GDP growth series. We also present the prediction and 95% confidence interval for 2015 Q4 at the foot of each specification we consider.

[4] Traditionally, forecasters have used macroeconomic variables to predict Chinese GDP growth. In a novel contribution, Nakamura, Steinsson and Liu (2016) have computed adjustments to official GDP figures using Engel curves derived from micro data consumption surveys in China, finding that GDP growth between 1995 and 2011 has been less smooth than implied by the official statistics.

5 Results

5.1 Specification

Figure 3 presents the national growth rates for official GDP and aggregate nighttime lights in China for the period 2005-2013. It is apparent that, in the aggregate time series, nighttime lights experiences erratic swings in growth, most noticeably in 2010. While that particular swing is caused by a new (and considerably more sensitive) nighttime lights recording satellite being launched in 2010, other swings cannot be traced to any specific satellite replacements, but rather should be attributed to changing detection properties of existing satellites. In fact, in the pure time series, Chinese official GDP growth rates are uncorrelated with Chinese nighttime lights growth rates. Hence, we need to flexibly control for changing satellite detection properties in our empirical analysis, which we do by employing year fixed effects in all specifications. Thus, we are identifying our estimates exclusively from differential growth in nighttime lights and economic activity measures across Chinese provinces, and not from the national-level series. As we are interested in growth rates rather than levels, we always include province fixed effects in all specifications. The baseline specification that we estimate is

$$y_{i,t}^{L} = \alpha_i + \lambda_t + \sum_{k=1}^{K} b^K y_{i,t}^{K} + \varepsilon_{i,t}$$

if the candidate proxy measures are $y_{i,t}^{1}, y_{i,t}^{2}, ...y_{i,t}^{K}$.

An additional concern could be that changes in the nighttime lights detection technology produce differential effects in different places, in particular because places with large agglomerations of bright lights may experience top-coding that censors part of the increase in lights that takes place when a satellite becomes more sensitive. We can address this concern by allowing changes in nighttime lights detection technology to have effects that vary with a variable that proxies for the extent of lights agglomeration, such as initial urbanization level. Hence, we also present estimates in which year fixed effects are interacted with 2004 provincial urbanization rates. We find that the resulting estimates are similar to the more parsimonious province and year fixed effects specification, providing confidence that controlling for year fixed effects removes the spurious variation generated by changing nighttime lights detection technology. Our specification including urbanization by year fixed effects becomes

$$y_{i,t}^{L} = \alpha_i + \lambda_t + \mu_t u_i^{2004} + \sum_{k=1}^{K} b^K y_{i,t}^{K} + \varepsilon_{i,t}$$

where u_i^{2004} is the urbanization rate of province i in 2004.

5.2 Descriptive Statistics

Table I presents associations between nighttime lights and various measures of the Chinese economy. All specifications include province and year fixed effects, so the resulting correlations should be interpreted as taking place between growth rates. The first three columns of Panel I show that the components of the Li Keqiang index have a strong and positive individual association with nighttime lights growth. In particular, column 1 of panel 1 shows that a 1 percentage point (pp) increase in the growth rate of electricity production is associated with a 0.28 pp increase in the growth rate of aggregate nighttime lights. The association between nighttime lights growth and railroad freight growth is somewhat weaker (0.079) and the association between nighttime lights growth and bank loan growth is somewhat stronger (0.41), foreshadowing results to come.

Columns 4 through 12 of Table I present individual associations between nighttime lights growth and the growth rates of the macroeconomic indicators that are used by the Bloomberg, CAP and Fernald indices described in Section 3.2, as well as the growth rate of our approximation of the Lombard Street indicator at the province by year level. Contrary to what we observe for the Li Keqiang variables, the associations between nighttime lights growth and the growth rates of these indicators are typically statistically insignificant (except for airport throughput, fixed asset investment and steel production in columns 8, 11 and 12). For example, the association between nighttime lights growth and completed floor space growth is a statistically insignificant 0.029 pp (panel 1, column 6), while the association between nighttime lights growth and rail passenger growth is negative. Hence, the Li Keqiang indicators appear to be considerably better suited to predicting Chinese economic activity than plausible alternative indicators.

Panel 2 of Table I replicates the specifications of Panel 1, but now adds in log real provincial GDP as an additional covariate. We see that the growth rate of real provincial GDP is significantly partially correlated with nighttime lights growth in all specifications (hence, for each of the other determinants of Chinese growth considered), except for the specification with bank loans (column 3). This observation shows that nighttime lights growth is tightly correlated with real provincial GDP growth in a panel setting (in contrast to the lack of correlation between nighttime lights growth and official GDP growth in the pure time series in Figure 3), but there is suggestive evidence that bank loan growth might be a potentially stronger predictor of Chinese economic activity. Panels 3 and 4 repeat the specifications of Panel 1 and 2, but now include year fixed effects interacted with 2004 urbanization rates in order to capture the potential heterogeneity of the impact of satellite changes on lights measurement in different Chinese provinces. We see that the estimated correlations are little changed, suggesting that year fixed effects are satisfactory in extracting the variation attributable to erratic satellite measurement quality.

5.3 Baseline Results

Table II presents the core results of the paper: the regressions of nighttime lights on the Li Keqiang variables. As noted in Section 3.2, there is a large discrepancy between official real GDP growth rates and the real growth rates implied by aggregating the provincial growth rates (Figure 4), but there is much less discrepancy between official nominal GDP growth rates and the nominal growth rates implied by provincial data. Hence, official real GDP growth apparently can be approximated as aggregated nominal provincial GDP growth deflated by a nationwide deflator. Therefore, we use log nominal provincial GDP as our measure of GDP (we investigate robustness to using real GDP later). Once again, we present results in two panels: specifications in the first have province and year fixed effects, while specifications in the second have province and year fixed effects interacted with provincial urbanization in 2004. At the foot of each regression, we report the prediction of Chinese 12-month growth in 2015 Q4 and its 95% confidence interval based on the regression and on the forecasting procedure described in Section 4.

Column 1 of Panel 1 presents the association between aggregate lights growth and nominal GDP growth in the panel of provinces. We see that a 1 pp increase in the growth rate of nominal GDP is associated with a 0.54 pp increase in the growth rate of aggregate nighttime lights. Hence, GDP growth and lights growth are strongly correlated, which is consistent with both of them being correlated with unobserved true income growth (recall that Assumption A1 requires lights growth to be correlated with true income growth). The forecast of growth in this column is just the official GDP growth rate in 2015 Q4, which was 6.6% in log points (whenever we try to forecast GDP with only a single variable, there are no standard errors, because a single slope coefficient is not identified in the mathematical framework of Section 2). Column 2 looks at the correlation between growth in nighttime lights and the Li Keqiang index (the average of the growth of electricity production, railroad freight and bank loans).[5] There is a statistically significant but much smaller association, with a 1 pp growth in the Li Keqiang being associated with a 0.086 pp growth in aggregate lights. The Li Keqiang index (normalized by regression to official GDP growth between 2005 and 2012) suggests that growth in 2015 Q4 was 6.3%, very slightly lower than the official figure. Column 3 includes both GDP and the Li Keqiang on the right hand-side of the regression. The coefficient on GDP declines in magnitude to 0.327 pp, while the coefficient on the Li Keqiang index does not change much. The predicted growth rate of official real GDP in 2015 Q4 is 5.9%, with a 95% confidence interval from 5.6% to 6.4%, and is statistically significantly lower than the official real GDP growth rate of 6.6%, though not by much, and

[5] We define the Li Keqiang variable as

$$(1/3) * \left(\ln\left(\text{electricity}_{i,t}\right) + \ln\left(\text{railfreight}_{i,t}\right) + \ln\left(\text{loans}_{i,t}\right)\right)$$

and when province fixed effects are included in our regressions, we are essentially looking at differences of this variable over time.

certainly not in a way that would indicate an alarming slowdown of Chinese growth. Figure 7 presents a time plot of the predicted growth rate of official real GDP from this regression. This plot looks very similar to the plot of the official real GDP growth rate, except in 2015 and early 2016, when it temporarily (but statistically significantly) dips below the official numbers.

Column 4 presents a regression of log aggregate lights on the logs of the three Li Keqiang measures included as separate variables. Recall that under our assumptions in Section 2, regressing nighttime lights on these variables would yield coefficients that are proportional to these variables' weights in the best unbiased linear predictor of unobserved true income. Therefore, if the original Li Keqiang index were optimal, we would expect each of the separate variables to get the same coefficient. Instead, a 1 pp increase in electricity growth is associated with a 0.153 pp increase in nighttime lights growth, a 1 pp increase in railroad freight growth with a 0.035 pp increase in nighttime lights growth, and a 1 pp increase in loan growth with a 0.275 pp increase in nighttime lights growth. These partial elasticities imply that an "optimized version" of the Li Keqiang index should place 33% of the weight on electricity growth, 8% of the weight on freight growth, and 59% of the weight on bank loans growth, instead of giving each measure a weight of $\frac{1}{3}$. A test of equality of the coefficients, reported below the regression, strongly rejects the null hypothesis. This rebalancing of the weight on the Li Keqiang index implies a predicted Chinese growth rate that is considerably *higher* than the one reported: the point prediction is 8.2%, with a 95% confidence interval between 7.1% and 9.2%.

Figure 10 part 1 plots our entire predicted series with the associated pointwise confidence interval (recalling from Section 4 that the predictions before 2012 are in-sample, while the predictions after 2012 are out-of-sample). We see that our methodology predicts Chinese GDP growth to have been lower than offiicial estimates before the crisis of 2008, to have experienced a shallower decline in 2008 and a stronger recovery in 2009 and 2010, and to have stabilized at a higher level after 2011. Across the 11-year period under consideration, our predicted Chinese growth series appears to be on a horizontal trend, while the official growth estimates are on a declining trend. Our predictions for past Chinese growth agree, at least in spirit, with some market participants' estimates. For example, our index agrees with both the Barclays and the Bloomberg indices in estimating Chinese growth to have been around 11% in 2007, and not to have experienced a substantial spike before the financial crisis, while the official GDP series records growth as spiking from 12% to 15% between 2005 and 2007. However, our predictions for current Chinese growth rates are above those of market participants, as well as of the Chinese government itself.

The explanation for why we get this contrarian prediction is that nighttime lights growth appears to be strongly correlated with bank loan growth net of national trends in both variables, and bank loan growth has not been declining in China as a whole, while railroad freight growth, which declined precipitously in 2015, is not very correlated with nighttime lights growth. Figure 2 shows that bank loan growth has not experienced the fall-off observed in railroad freight growth (and to a lesser extent, electricity production growth) since 2011. Hence, the prediction implied by the nighttime lights data and Assumptions A0-A2 is that Chinese true income growth did not fall off by very much either since 2010. It is worth noting that the flatness of our predicted time series for Chinese growth in Figure 8 closely resembles the shape of the bank loan growth series in Figure 2. It is also worth noting that the confidence interval for our latest predictions is considerable, ranging from close to the official estimate (slightly above, or slightly below) to an upper bound of slightly below 10% in late 2016. Our results suggest that true Chinese GDP growth is not appreciably lower than reported in the official statistics, and possibly may be higher.

Column 5 adds log nominal GDP to the three Li Keqiang measures on the right hand-side of the regression. The coefficient on nominal GDP declines further relative to the specification in column 3, and is now significant only at 10%. It is, however, still large relative to the coefficients on electricity and railroad freight, and is only slightly smaller than the coefficient on bank loans. We still reject the null that the coefficients on the three components of the Li Keqiang index are equal. The predicted official real GDP growth rate in 2015 Q4 is now 7.2%, just a little above the official rate of 6.6%, with a 95% confidence interval stretching from 6.7% to 9.2%. The path of the prediction over time (Figure 9) is somewhat more similar to the path of the official GDP growth rate than the one based on the regression in column 4; however, the nighttime lights still suggest that the growth of the Chinese economy was considerably closer to its pre-2008 level than the official statistics suggest.

As we have mentioned, we use nominal GDP as a right hand-side variable in Column 5 because official nominal GDP growth coincides much closer with the growth of the provincial aggregate nominal GDP than official real GDP growth coincides with the growth of the provincial aggregate real GDP. It is useful to check that our methodological choice does not particularly affect our results. Column 6 replicates column 5 with real GDP standing in for nominal GDP, and our coefficient estimates and growth rate predictions remain much the same.

One concern with our analysis may be that nighttime lights systematically reflect certain sectors of the economy to a greater extent than others, and that this may cause nighttime lights to place a greater weight on measures reflecting these sectors than might be warranted for measuring overall economic activity. For example, if most bank loans are funding construction projects and construction generates more nighttime light per unit of economic activity than other sectors, nighttime lights may assign too high a weight to bank loans. To address this concern as a first pass, we include the logs of primary, secondary and tertiary output as right hand-side variables in column 7 (the coefficients of these variables being omitted). We see that the coefficient on bank loans is unaffected, while the coefficient on electricity drops slightly, and predicted Chinese growth in 2015 Q4 rises to 8.2%.

A concern with the analysis is that Assumption A2 may fail for electricity, as measurement error in electricity may be positively correlated with measurement error in nighttime lights. To address this concern, we use the methodology of Section 2.2 to conservatively correct for the bias in electricity. Columns 8 and 9 perform this check under the assumptions that the covariance of electricity and lights is no smaller than the covariance between electricity and loans (which reduces the covariance of electricity and lights by a factor of 1.3; column 8) and that the covariance of the signal in electricity and lights is zero, so that their entire measured covariance is a result of correlated errors (column 9). While the coefficient on electricity shrinks (and becomes sharply negative in column 9), the coefficient on loans rises and remains significant, substantially *increasing* the prediction for the Chinese GDP growth rate in 2015 Q4 (albeit with similar confidence intervals to previous specifications). The truth is likely to be somewhere in between columns 5 and columns 9 (perhaps at column 8), but for any intermediate assumption between these two cases, the values of the regression coefficients will be intermediate to the ones in these two columns.

It is important to examine how our results vary for different regions of China and for different subperiods of our sample. China is a very economically diverse country, with most economic activity taking place along the Pacific coast, while the western part of the country is sparsely populated and thinly developed. A concern could be that our estimates are not representative of the economically most relevant regions of China. Moreover, China has been undergoing a structural transformation away from manufacturing and towards services, so different variables could, in principle, be better predictors of true Chinese growth rates at the end of our sample period than at the beginning. Table III checks the robustness of our estimates to excluding different subsamples of years and Chinese regions. Column 1 replicates the baseline for reference. Columns 2 through 4 check robustness to exclusion of one of the three major groups of provinces (coastal, central and western). Columns 5 through 7 restrict the sample to the period prior to (and including 2008), the period following (and including) 2010, and the entire period excluding the crisis years 2008-2010, respectively. We see that in all robustness checks, the coefficient on loans is remarkably stable and similar to the baseline. While the magnitudes of the other coefficients vary, they remain qualitatively similar to the baseline, except for the coefficient on log GDP. The predicted growth rate of the Chinese economy in 2015 Q4 never falls below the baseline estimate, and the estimated 95% confidence intervals of the growth rate predictions are remarkably stable across specifications.

In both Table II and Table III, Panel 2 presents results from all the above regressions, but now including year by 2004 provincial urbanization fixed effects. These more flexible fixed effects allow the effects of changing technology of lights measurement to be heterogeneous by province as a function of urbanization. Despite the added flexibility of the model, the results are much the same as in Panel 1. We reliably find that our best estimate of the true unobserved growth rate of the Chinese economy in 2015 Q4 is inconsistent with a sharp slowdown of the Chinese economy at the time, and may be consistent with higher Chinese growth than in the official statistics.

6 Results for Other Indicators

6.1 Baseline Estimates

Besides the Li Keqiang index, there exist many other efforts of market participants to understand possible departures of the Chinese economy from the official statistics, which were discussed in Section 1. We consider the following four indices: the Barclays GDP tracker, the Bloomberg monthly GDP estimate,[6] the China Activity Proxy (CAP) produced by Capital Economics, and the Lombard Street Research real GDP indicator. In tables IV through V we perform a similar analysis of these indices as we do for the Li Keqiang variables, and we check the robustness of our earlier conclusions about the Li Keqiang variables when the variables included in alternative indices of the Chinese economy are controlled for.We know the predictors that are combined by the Barclays index to generate its estimates of Chinese growth rates, but cannot use several of them (3 variables relating to purchasing managers' indices as well as auto sales) as they are not defined at the provincial level for China. We do not know the formulas generating the Bloomberg and the CAP indices, but we believe that they are linear functions of the variables that they use, so we include these variables in our regressions. We use an approximation to the formula for the Lombard Street index, which reweights nominal output from various sectors using different price indices.

In addition to indices created by forecasters in the financial industry, we also consider the methodology of John Fernald et al. (2015), who extracts principal components from a broad set of potential growth determinants, and then regresses data on exports to China from developed countries on the resulting series. Fernald's argument is that exports to China are measured by developed countries, and therefore, their errors should be uncorrelated with measurement errors in economic variables reported by China. We prefer to analyze Fernald's approach by including all the variables that he uses that are available at the provincial level rather than extracting their principal component because including the variables separately induces the regression to assign them coefficients that maximize the explanatory power of the regression, and therefore, that can shift the coefficients on variables that we are already considering.

Table IV presents regressions of log aggregate nighttime lights on the components of the Li Keqiang, Barclays, Bloomberg, CAP and Fernald indices, as well as on our approximation to the Lombard Street index. We refer to all indices as "pseudo" indices to highlight that we are not analyzing the exact indices

[6] Bloomberg also considers a modified Li Keqiang index in which bank loans and electricity each get 40% of the weight, while railroad freight gets 20% of the weight. Based on our coefficient estimates in Table II column 5, we can reject this pattern of weighting at the 3% significance level.

used by these researchers, but rather some of their components or our approximations of them. Here, the regressions using year by urbanization fixed effects are side by side, rather than in separate panels. Column 1 reproduces Column 4 in Table II and reproduces the regression of log nighttime lights on the separated components of the Li Keqiang index. Column 2 reports estimates of the proxy depending on the available variables of the Barclays index. We see that electricity, FAI and retail sales receive considerable weight. Column 3 reports estimates for the proxy containing variables in the Bloomberg index; its components are a subset of the Barclays index components, and once again, electricity, FAI and retail sales receive most of the weight. Column 4 reports estimates for the proxy only depending on the Lombard Street indicator evaluated at the provincial level (with national-level price indices for imports and exports). We note that the Lombard Street indicator is just log nominal GDP minus the log of a weighted average of several price deflators (the CPI, the FAI price index, and the export and import price indices). Column 5 reports the regression of log nighttime lights on the components of the CAP index. It is clear that electricity is helpful in predicting economic activity, while the total freight, passenger and the floor space completed variables aren't. Lastly, Column 6 reports the regression of log nighttime lights on eight of the variables analyzed by Fernald et al. (2015). Once again, the coefficients on the three Li Keqiang variables do not change with the addition of the other indicators.

The different indices offer divergent predictions for Chinese GDP growth in 2015 Q4. At one extreme, using the Li Keqiang covariates suggests that Chinese growth in 2015 Q4 was 8.2%. The pseudo CAP, pseudo Lombard Street and pseudo Fernald indices suggest growth rates relatively consistent with the official data at 7.1%, 6.8% and 6.6% respectively.[7] On the other hand, the point estimate prediction using the Barclays covariates is 4.3%, with a lower confidence bound of 3.8%, which would be consistent with a considerable slowdown in China's growth rate, though is still above the Lombard Street official calculation of under 3%. The pseudo Bloomberg index lies in between, with a prediction that the Chinese economy grew by 5.3% in 2015 Q4, but a lower confidence bound of 4.7%. However, the upper confidence bounds on the predictions based on these specifications are considerably above the officially reported 2015 Q4 growth rate of 6.6% (they are 7.9% for the pseudo Barclays and 8.3% for the pseudo Bloomberg) suggesting that these specifications could still be compatible with the Chinese economy growing at officially reported rates or higher. Columns 7 through 12 of Table IV present results for specifications with 2004 urbanization by year fixed effects; the

[7] The fact that our prediction for Chinese GDP growth in 2015 Q4 using the Lombard Street indicator is 6.8% is very surprising because the predicted growth rate of Chinese GDP in 2015 Q4 according to the methodology of Lombard Street is 2.9%. The reason for this discrepancy lies in the fact that our methodology normalizes any predictor of Chinese GDP growth by regressing on it the official Chinese GDP growth series between 2005 and 2012. Hence, if that series reports systematically lower or higher growth rates, or if it reports systematically more or less volatile growth rates, the prediction regression will remove this source of discrepancy of the series with GDP. It turns out that the Lombard Street index can indeed be seen as being very close to a linear transformation of official Chinese growth rates (it has larger booms and larger busts, but the booms and the busts take place roughly contemporaneously with the ones in the official series), so our prediction methodology undoes this transformation, and the Lombard Street index turns out to be very close to official GDP.

predictions are slightly higher for the pseudo Barclays and pseudo Bloomberg indices and slightly lower for the other indices, but not substantively different. Figures 10 and 11 (top panels) present graphs of the predicted paths of Chinese growth rates from columns 1-6 implied by the different proxies in combination with our assumptions. We see that the Barclays and the Bloomberg indices imply a slowdown that started after 2013 and reached its nadir around the end of 2015, while the other indices imply a much gentler downward trend in Chinese GDP growth that lies above the official estimates.

Table V presents the same regressions as Table IV, but now also includes log provincial GDP as a control. These regressions are similar to Column 5 in Table II. The coefficient on log provincial GDP is not statistically significant in specifications involving the pseudo Barclays or the pseudo Fernald indices, but tends to be significant at least at the 10% level in the others. It is worth noting that in specifications including the Lombard Street index, the coefficient on log GDP shoots up in order to offset the log GDP term inside the Lombard Street index, and that the negative coefficient on the Lombard Street index should be interpreted as a positive coefficient on the log of the Lombard Street deflator. The predicted Chinese growth rates for 2015 Q4 are also, generally, closer to the officially reported value of 6.6%, but the predictions using the pseudo Barclays index (4.7% in column 2) is still compatible with a considerable slowdown in Chinese growth, as is the lower confidence bound on the prediction using the pseudo Bloomberg index (4.4% in column 3).

6.2 Adding the Li Keqiang Variables

From Tables IV and V we learn that two of the indices that do not include the bank loans measure as a component – the pseudo Bloomberg and the pseudo Barclays – suggest predictions for recent growth rates of the Chinese economy that are markedly lower than those obtained when bank loans are used as a predictor, as well as the officially reported growth rates. It is therefore important to examine the predictions for Chinese growth rates when both the bank loans measure (and the Li Keqiang variables more generally) as well as the pseudo Bloomberg and pseudo Barclays variables (nominal retail sales, nominal fixed asset investment as well as others) are included together on the right hand-side of regression (1). We consider such specifications in Table VI, which presents regressions of log nighttime lights on log GDP, the three Li Keqiang variables and the variables from other indicators of the Chinese economy. Additionally, these regressions can be seen as tests of the exclusion restriction in the Li Keqiang regressions of Table II by analyzing their robustness to inclusion of variables that other researchers have claimed to be important for predicting Chinese economic growth. Once again, columns 1 through 6 present results from regressions with province and year fixed effects, while columns 7 through 12 present results with province and 2004 urbanization by year fixed effects.

Column 1 of Table VI replicates column 5 of Table II and presents our baseline result that using the Li Keqiang variables separately alongside with log nominal GDP suggests a growth rate for China in 2015 Q4 that is slightly higher than the official measure (7.2%). Column 2 now adds the variables from the pseudo Barclays index to this regression. We see that the coefficients on the Li Keqiang variables remain remarkably stable, with the coefficient on the loans variable declining slightly from 0.224 in Column 1 to 0.206 in Column 2, but retaining significance at the 5% level. In contrast, the coefficients on electricity and retail sales change considerably relative to those in the second columns of Tables IV and V. The prediction for Chinese growth in 2015 Q4 rises to 5.6%, which is higher than the predictions using the Barclays, CAP and Lombard Street indices discussed in Section 3, albeit with a lower confidence bound of 4.7%, which is close to the official prediction of the CAP index. Column 3 of Table VI replicates our analysis using the covariates from the pseudo Bloomberg index (which are a subset of the covariates in the pseudo Barclays index). We observe results similar to those in the previous column. The coefficient on bank loans remains statistically significant and similar to our baseline result, while the coefficients on the components of the pseudo Bloomberg index vary more markedly, and the prediction for Chinese economic growth in 2015 Q4 rises relative to the prediction in Column 3 of Tables IV and V (from 5.3% to 5.8%), though its lower confidence bound (5.1%) is still compatible with considerably lower (though not catastrophically lower) rates of Chinese growth than reported in the official statistics. Columns 4 through 6 investigate the effect of adding the full suite of Li Keqiang variables (and log nominal GDP) to the pseudo Lombard, pseudo CAP and pseudo Fernald indices (the latter already included all the Li Keqiang variables). We observe that for all of them, the coefficient on bank loans remains remarkably stable and close to the baseline, while predicted Chinese growth rate in 2015 Q4 is close to the officially reported growth rate of 6.6%. Our results change little when we add 2004 urbanization by year fixed effects in columns 7 through 12.

We can therefore conclude that while using only the variables in the pseudo Barclays and pseudo Bloomberg indices can generate low predictions for Chinese growth in 2015 Q4 (as well as more generally), it is appropriate to supplement these variables with additional variables out of the Li Keqiang index, especially with the bank loans variable, and once this is done, the predicted growth rates rise to levels that are not severely inconsistent with officially reported growth rates, and that are incompatible with the most pessimistic views of the Chinese economy at that time. Moreover, the confidence intervals on the predictions for all specifications allow for the possibility of Chinese growth being considerably higher than the officially reported growth rate. The subsequent subsection will show that further improvement of the set of predictors leads us further towards a more optimistic view of Chinese growth.

6.3 Adding Deflators

So far, the monetary variables that we have used as predictors of log nighttime lights in provincial regressions have all been nominal, in order to be consistent with our use of nominal GDP as well as with the way that these variables are used in the indices that we have considered. However, our framework allows us to evaluate whether or not these variables should be deflated and in what way simply by adding the logarithms of candidate deflators as right hand-side covariates to the regression (1). The resulting model nests our old model as a special case, allowing us to understand whether the deflators may have been important omitted variables in the earlier model. We employ three deflators: 1) the retail price index, which we include in all specifications containing retail sales, 2) the FAI price index, which we include in all specifications containing bank loans or real estate investment, and 3) the GDP deflator, which we include in all specifications containing GDP.

Table VII replicates Table IV, but now accounts for the deflators. Column 1 presents results for the Li Keqiang variables. We see that the coefficients on the Li Keqiang variables change little, while the log FAI price deflator enters with a large, positive, but insignificant coefficient. The positive coefficient on the log FAI price deflator implies that instead of dividing bank loans by the price deflator, we should be multiplying bank loans by (a function of) the deflator, suggesting that Chinese regions with higher growth of investment prices also experience higher nighttime lights growth (and hence, higher unobserved true income growth). Therefore, investment prices may be measuring not just the price of investment goods, but also unobservables associated with growth. The predicted 2015 Q4 Chinese growth rate is 7.7%, somewhat less than in Table IV, but still higher than the officially reported growth rate.

Columns 2 and 3 present results for the components of the pseudo Barclays and pseudo Bloomberg indices respectively. Recall that in Table IV we observed that using these components as right hand-side variables without any deflators generated predictions for Chinese growth that were considerably lower than the officially reported growth rates. Now, once FAI and retail price index deflators have been included, these predictions are around 8.8%, comfortably above the officially reported growth rates. While the coefficient on the log FAI price index is small, the coefficient on the retail price index is large (around unity) and statistically significant at the 10% level. Since both of these specifications nested the specifications in Columns 2 and 3 of Table IV, we become more confident that Chinese growth in 2015 Q4 was relatively high because having given the data a choice between models with different predictions for Chinese growth, the data chose the model predicting higher growth.

Columns 4 through 6 of Table VII present results for the components of the pseudo Lombard, pseudo CAP and pseudo Fernald indices. Neither the pseudo Lombard nor the pseudo CAP indices require the addition of deflators (the pseudo Lombard index provides its own deflator, while the pseudo CAP index uses only intrinsically real quantities, such as freight volume or passenger traffic), so these columns are identical to corresponding columns in Table IV. When we consider the pseudo Fernald index, we observe that the weight on the log FAI price index is tiny, while the weight on the log retail price index is 1.27 and statistically significant at 1%, with the predicted growth rate in 2015 Q4 rising to 9.4% (relative to 6.6% in column 6 of Table IV). Adding deflators to this specification, if anything, reinforces our result that Chinese growth in 2015 Q4 was not substantially lower than officially reported, and may have even been higher. Columns 7 through 12 of Table VII present results of the same regressions as columns 1 through 6, but now including 2004 urbanization by year fixed effects, with the results being little changed.

Figures 10 and 11 (bottom panels) present the full time paths and confidence intervals of our predicted Chinese growth rates from the specifications estimated in columns 1 through 6 of Table VII. We observe that the resulting paths of GDP growth rates are considerably flatter and closer to Chinese long-run average growth than the ones coming from specifications without deflators in Table IV (top panels). However, we note that the confidence intervals in the last few years include growth rates that are essentially equal to the officially reported ones as well as growth rates substantially above that level.

Table VIII replicates Table VII but now controls for log real GDP and the log of the GDP deflator in each specification. Note that log nominal GDP is a linear combination of these two variables, so these specifications are agnostic about whether we should use nominal or real GDP. Our conclusion that Chinese growth was not substantially below, and could have been above the officially reported growth rates in 2015 Q4 remains valid, with all the point estimates of predicted Chinese growth rates in 2015 Q4 no smaller than 6.4%. The only potential concern is that in the pseudo Barclays specification without urbanization by year fixed effects (column 2), the lower confidence bound for the prediction is 4.6%, which is in line with the official estimate of the CAP index (though still above that of the Lombard Street indicator). The coefficient on real GDP is statistically significant at least at the 10% level in all the specifications excluding log bank loans – specifically, all except the Li Keqiang (column 1) and the pseudo Fernald (column 6).

Table VIII replicates Table VI in including all the Li Keqiang components in every specifications, but now controls for the relevant deflator variables. We see that once again, the point estimates of our predictions of Chinese growth in 2015 Q4 are close to or above the officially reported growth rate, with the lowest one being 6.4% (for the pseudo CAP with urbanization by year fixed effects). The lowest value of the lower confidence bound is 5.2%, slightly below the official Barclays prediction. In all specifications, the coefficient on the log bank loan volume variable is significant at least at the 10% level and ranges from 0.19 to 0.28 in magnitude. Since there is some heterogeneity in the confidence bounds on the predictions of the various indicators, we run a specification including all the macroeconomic proxies from all the indices that we have considered, together with their deflators, in column 7. We predict Chinese growth in 2015 Q4 to have been 8.6%, with a 95% confidence bound from 7% to 10.6%. (The results in column 14, which include 2004 urbanization by year fixed effects, are similar, except the lower confidence bound extends to 6.5%). While this specification is very large and raises concerns of overfitting, we find it reassuring that when we allow the best unbiased linear estimator to depend on every macroeconomic proxy used in all the indices in the analysis, our predictions are remarkably in line with our baseline results, suggesting that Chinese growth in 2015 Q4 was not substantially lower than officially reported, and may have been considerably higher. Figure 12 presents the graph of the resulting time path of predicted Chinese growth rates; we observe that it is similar to the paths of the pseudo Barclays, Bloomberg and Fernald indices with deflators included, being shallower than the official growth rate series, and in particular, suggesting a more gentle decline in growth rates in the 2010s.

7 Conclusion

There is a suspicion among analysts and the popular press at Chinese official statistics have overstated recent growth. We provide a novel approach based on our recent research (Pinkovskiy and Sala-i-Martin 2016a and b) that uses satellite-recorded nighttime lights data to compute the optimal way to weight macroeconomic proxies (including GDP) in predicting unobserved true income and its growth rate. In this paper, we use this methodology to predict Chinese growth using proxy variables frequently cited in the literature (in particular, the variables comprising the Li Keqiang index among others). We find that the Li Keqiang variables perform well in predicting nighttime lights growth in China, although we reject the equal weighting of the variables, finding instead that bank loans should get considerably more weight while railroad freight should get considerably less. Hence, according to our assumption, the Li Keqiang variables with modified weights do a good job at predicting true unobserved Chinese economic growth. This result changes little when we add other variables used by the literature to forecast Chinese GDP.

Using our estimates, we can construct predictions of Chinese growth rates and compare them to the official statistics and to the predictions of various forecasters active in the literature. Our results show that Chinese growth rates are not appreciably less than officially reported growth rates, and may be considerably above them, in a way consistent with no sharp slowdown in Chinese growth over the 2014-2016 period. This finding is surprising because most other indicators trying to calculate Chinese growth estimate an even sharper decline in growth than is present in the official statistics.

The intuition for our finding that Chinese growth is no lower than in the official statistics is that nighttime lights growth across Chinese provinces (netting out changes that are common at the national level) is more correlated with growth indicators that have remained roughly stable (bank lending, and, to a lesser degree, electricity) than with indicators that have declined (such as railroad freight). However, our result does not depend on including bank lending as one of the predictors. Given a sufficient breadth of possible right hand-side variables in a regression, the nighttime lights tend to select the ones consistent with a no less optimistic growth performance relative to trend than is present in the official real GDP growth series.

Why might Chinese official statistics be understating, rather than overstating, GDP growth? While our methodology allows us to obtain this finding, it does not shed light on the mechanisms behind it. One explanation consistent with our results could be if Chinese national accounts understate the growth rate of services, which would progressively understate growth as services become a larger fraction of the Chinese economy (for example, Rosen and Bao (2015) argue that Chinese national accounts data underestimate the level of the services sector, especially of its real estate component). More generally, there have been concerns that GDP growth is understated in many developed countries as well because of the failure to account for structural transformations towards new goods and services, especially those delivered electronically (Boskin Commission 1996). Investigating both of these potential explanations quantitatively remains an open question for future research.

References

[1] **Adams, F. Gerard, and Yimin Chen.** "Skepticism about Chinese GDP growth–the Chinese GDP elasticity of energy consumption." *Journal of Economic and Social Measurement* 22, no. 4 (1996): 231-240.

[2] **Boskin, Michael J.** Toward a more accurate measure of the cost of living: Final report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. Advisory Commission to Study the Consumer Price Index. 1996

[3] **Chen, Xi, and William D. Nordhaus**. "Using Luminosity Data as a Proxy for Economic Statistics" Proceedings of the National Academy of Sciences. 2011

[4] **Doll, Christopher N.H.** "CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications." Manuscript. 2008

[5] **Fernald, John, Eric Hsu, Mark M. Spiegel.** 2015. "Is China Fudging its Figures? Evidence from Trading Partner Data." *Federal Reserve Bank of San Francisco Working Paper* 2015-12.

[6] **Henderson, J. Vernon, Adam Storeygard, and David N. Weil**. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2012), no. 2: 994-1028.

[7] **Holz, Carsten A.** "The quality of China's GDP statistics." *China Economic Review* 30 (2014): 309-338.

[8] **Maddison, Angus, and Harry Wu**. "China's Economic Performance: How Fast has GDP Grown? How Big is it Compared with the USA?." University of Queensland (2006).

[9] **Michalopoulos, Stelios, and Elias Papaioannou**. "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81 (2013), no. 1: 113-152.

[10] **Michalopoulos, Stelios, and Elias Papaioannou**. "National Institutions and Sub-national Development in Africa." *Quarterly Journal of Economics* 129 (2014) no. 1:151-213.

[11] **Nakamura, Emi, Jón Steinsson, and Miao Liu.** "Are Chinese Growth and Inflation too Smooth? Evidence from Engel Curves." *American Economic Journal: Macroeconomics* 8, no. 3 (2016): 113-144.

[12] **National Geophysical Data Center (NGDC)**. "Version 4 DMSP-OLS Nighttime Lights Time Series." National Oceanic and Atmospheric Administration. http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html. 2010. (accessed September 2013).

[13] **Orlik, Tom and Fielding Chan**. 2015. "China Proxy Wars." *Bloomberg Brief: Economics Asia*.

[14] **Perkins, Dwight H., and Thomas G. Rawski**. "Forecasting China's economic growth to 2025." *China's great economic transformation* (2008): 829-86.

[15] **Pinkovskiy, Maxim, and Xavier Sala-i-Martin.** 2016a. "Lights, Camera... Income! Illuminating the National Accounts-Household Surveys Debate." *Quarterly Journal of Economics* 131, no. 2: 579-631.

[16] **Pinkovskiy, Maxim, and Xavier Sala-i-Martin.** 2016b. "Newer need not be better: evaluating the Penn World Tables and the World Development Indicators using nighttime lights." *National Bureau of Economic Research* No. w22216.

[17] **Rawski, Thomas G.** "What is happening to China's GDP statistics?." *China Economic Review* 12, no. 4 (2001): 347-354.

[18] **Rosen, Daniel H. and Beibei Bao.** 2015. *Broken Abacus? A More Accurate Gauge of the Chinese Economy.* CSIS.

8 Tables

Table I

China: Possible Indicators of Growth																
Dependent Variable is Log Light Intensity																
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Elect Ricity	Rail Freight	Bank Loans	Total Freight	Retail Sales	VAI	FAI	Exports	Imports	Steel Prod.	LS GDP	Floor Sp. Under Const.	Passen Gers	RE Investment	Air Thrput	Floor Starts
Province and Year Fixed Effects																
Log Ind. Var.	.282***	.068**	.411***	.056	.811***	.165	.227***	.042	-.007	.078***	.482***	.094	-.005	.053	.214***	.062**
	(.044)	(.031)	(.102)	(.037)	(.293)	(.118)	(.060)	(.032)	(.033)	(.027)	(.135)	(.064)	(.025)	(.044)	(.058)	(.028)
Within R2	.13	.05	.14	.01	.06	.02	.08	.01	.00	.02	.07	.02	.00	.02	.08	.03
GDP and Province and Year Fixed Effects																
Log GDP	.358***	.474***	.311**	.552***	.444***	.681***	.416***	.530***	.543***	.510***	1.853***	.506***	.543***	.523***	.397**	.499***
	(.119)	(.126)	(.123)	(.125)	(.145)	(.177)	(.137)	(.120)	(.120)	(.125)	(.512)	(.123)	(.122)	(.137)	(.179)	(.125)
Log Ind. Var.	.209***	.040**	.308***	-.006	.408	-.145	.143**	.034	-.011	.049	-1.446***	.034	.003	.011	.114	.029
	(.034)	(.016)	(.092)	(.040)	(.295)	(.145)	(.071)	(.025)	(.027)	(.032)	(.551)	(.049)	(.022)	(.039)	(.079)	(.026)
Within R2	.17	.12	.17	.11	.12	.12	.14	.12	.11	.12	.15	.11	.11	.11	.13	.11
Province, Year and Urban Pct. by Year Fixed Effects																
Log Ind. Var.	.258***	.070**	.358***	.032	.487	.090	.193***	-.014	-.020	.071***	.402**	.127***	-.017	.058**	.184***	.065***
	(.052)	(.033)	(.116)	(.047)	(.465)	(.135)	(.056)	(.046)	(.033)	(.019)	(.170)	(.048)	(.027)	(.026)	(.062)	(.023)
Within R2	.10	.05	.11	.00	.01	.00	.06	.00	.00	.02	.05	.05	.00	.03	.07	.03
GDP and Province, Year and Urban Pct. by Year Fixed Effects																
Log GDP	.324***	.389***	.292**	.499***	.460***	.646***	.370**	.469***	.471***	.439***	1.791***	.377**	.466***	.427**	.342	.414***
	(.125)	(.147)	(.125)	(.147)	(.153)	(.199)	(.172)	(.153)	(.145)	(.153)	(.486)	(.176)	(.147)	(.175)	(.216)	(.157)
Log Ind. Var.	.198***	.047**	.269***	-.021	.055	-.192	.127*	-.001	-.022	.049*	-1.441***	.076	-.005	.024	.101	.038
	(.032)	(.020)	(.099)	(.046)	(.416)	(.152)	(.071)	(.041)	(.026)	(.026)	(.533)	(.052)	(.024)	(.031)	(.087)	(.026)
Within R2	.14	.11	.13	.09	.08	.10	.11	.08	.09	.09	.13	.10	.08	.09	.10	.10
Number of Obs.	270	270	270	270	270	270	270	270	270	269	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

(I)

Table I presents regressions of log light intensity on the independent variables in the column headings using the panel of Chinese provinces described in Section 3. Standard errors clustered on province in parentheses. All variables are from or based on the CEIC. The variables are 1) log electricity production, 2) log railroad freight volume, 3) log total nominal value of loans made by banks, 4) log total freight volume, 5) log nominal retail sales, 6) log value added of industry, 7) log fixed asset investment expenditure, 8) log nominal exports, 9) log nominal imports, 10) log steel production, 11) authors' estimate of provincial-level log GDP using the methodology of the Lombard Street indicator, 12) log floor space under construction, 13) log total passengers, 14) log real estate invesment, 15) log airport throughput, 16) log floor starts. More data on these variables is available in the Data Appendix.

Table II

China: Li Keqiang Regressions									
Dependent Variable is Log Light Intensity: Prediction is for 2015 Q4									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	GDP Only	Li Keqiang Only	GDP and LK	Separate Measures	Baseline	Real GDP Baseline	Industry Controls	Elect. Slope Like Loans	Electricity Zero Effect
Province and Year Fixed Effects									
Log GDP	.541***		.335**		.210*	.246	-.477	.245*	.362
	(.120)		(.141)		(.125)	(.179)	(.403)	(.145)	(.228)
Li Keqiang		.086***	.062***						
		(.026)	(.022)						
Log Electricity				.153***	.134***	.156***	.106**	.038	-.279*
				(.045)	(.043)	(.045)	(.042)	(.077)	(.165)
Log Freight				.035**	.027**	.030**	.029***	.033	.050
				(.013)	(.011)	(.012)	(.011)	(.028)	(.051)
Log Loans				.275***	.224***	.241***	.262***	.278**	.454**
				(.092)	(.086)	(.091)	(.090)	(.132)	(.230)
Within R2	.11	.13	.16	.19	.20	.20	.22	.20	.20
P-value 3 coeffs equal	0	0	0	0	0	0	0	.04	.04
Prediction	6.6	6.3	5.9	8.2	7.2	7.1	8.2	8.0	10.6
CI For Prediction	(6.6,6.6)	(6.3,6.3)	(5.6,6.4)	(7.1,9.2)	(6.7,9.2)	(6.6,9.4)	(6.7,10.2)	(6.4,10.5)	(7.5,10.9)
Province, Year and Urban Pct. by Year Fixed Effects									
Log GDP	.470***		.276*		.194	.243	-.152	.218	.297
	(.146)		(.150)		(.125)	(.176)	(.439)	(.163)	(.246)
Li Keqiang		.080***	.062***						
		(.024)	(.021)						
Log Electricity				.144***	.129***	.145***	.111***	.046	-.227
				(.043)	(.041)	(.045)	(.041)	(.079)	(.183)
Log Freight				.040**	.033**	.036**	.032**	.038	.055
				(.016)	(.014)	(.015)	(.013)	(.039)	(.069)
Log Loans				.231**	.187*	.200**	.223**	.230	.371*
				(.105)	(.100)	(.100)	(.098)	(.145)	(.220)
Within R2	.08	.12	.14	.16	.17	.17	.18	.17	.17
P-value 3 coeffs equal	0	0	0	.01	.01	.01	.01	.16	.1
Prediction	6.6	6.3	5.8	7.9	6.9	6.8	7.8	7.4	10.7
CI For Prediction	(6.6,6.6)	(6.3,6.3)	(5.6,6.4)	(6.9,9.2)	(6.5,9.1)	(6.4,9.1)	(6.6,10.0)	(5.9,10.6)	(7.3,11.3)
Number of Obs.	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27

(II)

Table II presents regressions of log light intensity on log provincial GDP and the components of the Li Keqiang index. Standard errors clustered on province in parentheses. All independent variables are from the CEIC; see the Data Appendix for details. All predictions computed using the method explained in Section 4. Column 7 controls for log nominal GDP in primary industry, log nominal GDP in secondary industry and log nominal GDP in tertiary industry; the coefficients on these variables are available on request. Column 8 provides estimates of the regression coefficients accounting for potential correlation between the measurement error in log light intensity and log electricity production under the assumption that the elasticity of log electricity production with respect to log unobserved true income is the same as the elasticity of log total value of loans made by banks with respect to log unobserved true income. Column 9 provides the same estimates under the assumption that the elasticity of log electricity production with respect to log unobserved true income is zero.

Table III

China: Li Keqiang Regressions, Subsamples							
Dependent Variable is Log Light Intensity: Prediction is for 2015 Q4							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	No West China	No Central China	No Coastal China	Before 2008	After 2010	Excluding Crisis
Province and Year Fixed Effects							
Log GDP	.210*	.269	.331***	-.003	.074	-.197	.159*
	(.125)	(.266)	(.109)	(.159)	(.269)	(.354)	(.095)
Log Electricity	.134***	.066	.143***	.103**	.040	-.017	.166***
	(.042)	(.095)	(.046)	(.041)	(.117)	(.052)	(.039)
Log Freight	.027**	.009	.014**	.041**	.022**	-.026	.011
	(.011)	(.079)	(.007)	(.019)	(.010)	(.040)	(.008)
Log Loans	.224***	.311*	.223**	.224***	.309***	.218*	.300***
	(.086)	(.181)	(.110)	(.083)	(.117)	(.117)	(.077)
Within R2	.20	.10	.33	.16	.11	.05	.44
P-value 3 coeffs equal	0	.24	0	0	.06	.19	0
Prediction	7.2	8.3	7.1	8.1	9.4	9.5	8.0
CI For Prediction	(6.7,9.2)	(6.3,10.6)	(6.6,9.0)	(6.9,10.5)	(6.7,10.5)	(5.8,10.8)	(7.1,9.4)
Province, Year and Urban Pct. by Year Fixed Effects							
Log GDP	.194	.265	.321***	-.008	.082	-.284	.158*
	(.125)	(.246)	(.081)	(.180)	(.278)	(.416)	(.092)
Log Electricity	.129***	.161	.107**	.117***	.048	-.024	.151***
	(.041)	(.119)	(.047)	(.042)	(.135)	(.049)	(.034)
Log Freight	.033**	-.040	.016	.044**	.028***	-.019	.016
	(.014)	(.093)	(.010)	(.017)	(.009)	(.041)	(.010)
Log Loans	.187*	.247	.251**	.181	.286**	.190	.290***
	(.100)	(.184)	(.098)	(.121)	(.127)	(.120)	(.081)
Within R2	.17	.10	.32	.16	.11	.04	.38
P-value 3 coeffs equal	.01	.16	0	.01	.13	.29	0
Prediction	6.9	8.3	7.3	7.6	9.0	9.0	7.9
CI For Prediction	(6.5,9.1)	(6.1,11.6)	(6.5,9.4)	(6.6,10.5)	(6.3,10.6)	(6.3,10.4)	(7.0,9.3)
Number of Obs.	270	160	190	190	135	108	189
Number of Clusters	27	16	19	19	27	27	27

(III)

Table III presents regressions of log light intensity on log provincial GDP and the components of the Li Keqiang index for selected subsamples. Standard errors clustered on province in parentheses. All independent variables are from the CEIC; see the Data Appendix for details. All predictions computed using the method explained in Section 4. The division of China into the three regions is shown in Figure 1. Crisis years are defined to be 2008-2010.

8.1 Results with Alternative Indices

Table IV

	China: Alternative Indices											
	No Urban Pct. by Year FE						*Urban Pct. by Year FE*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald
	Province, Year and Urban Pct. by Year Fixed Effects											
Log Electricity	.153***	.214***	.241***		.253***	.109**	.144***	.220***	.254***		.217***	.120**
	(.045)	(.042)	(.034)		(.046)	(.054)	(.043)	(.042)	(.043)		(.048)	(.054)
Log Rail Freight	.035**					.032**	.040**					.037***
	(.013)					(.014)	(.016)					(.013)
Log Loans	.275***					.210*	.231**					.196*
	(.092)					(.111)	(.105)					(.111)
Log Total Freight		-.022			.022			-.002			.020	
		(.037)			(.036)			(.037)			(.041)	
Log Retail Sales		.438**	.346**			.350		.352	.309			.291
		(.211)	(.170)			(.214)		(.302)	(.235)			(.256)
Log VAI		-.015	-.062					-.041	-.067			
		(.136)	(.117)					(.139)	(.116)			
Log FAI		.188***	.180***					.201***	.186***			
		(.059)	(.066)					(.063)	(.068)			
Log Exports		.017	.015			.027		.006	.008			.024
		(.029)	(.032)			(.029)		(.043)	(.044)			(.040)
Log Imports		-.038						-.041				
		(.030)						(.030)				
Log Steel Production		.045						.038				
		(.030)						(.033)				
Log LS GDP				.482***						.402**		
				(.135)						(.170)		
Log Floor Space					.060						.092**	
					(.045)						(.042)	
Log Passengers					-.002						-.002	
					(.024)						(.022)	
Log RE Investment						.022						.032
						(.026)						(.027)
Log Air Thrput						.093*						.074
						(.053)						(.047)
Log Floor Starts						.006						.005
						(.017)						(.026)
Within R2	.19	.22	.20	.07	.14	.24	.16	.18	.17	.05	.13	.19
Prediction	8.2	4.3	5.3	6.8	7.1	6.6	7.9	4.4	5.2	6.8	6.8	6.3
CI For Prediction	(7.1,9.2)	(3.8,7.9)	(4.7,8.3)	(6.8,6.8)	(6.4,8.4)	(6.0,9.8)	(6.9,9.2)	(3.8,10.3)	(4.9,10.2)	(6.8,6.8)	(6.4,7.9)	(5.6,10.0)
Number of Obs.	270	270	270	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27

(IV)

Table IV presents regressions of log light intensity on components of the indices in the column headings. Standard errors clustered on province in parentheses. All predictions computed using the method explained in Section 4. All independent variables are from the CEIC and listed in the note to Table I; see the Data Appendix for details.

Table V

China: Alternative Indices and GDP												
	No Urban Pct. by Year FE						*Urban Pct. by Year FE*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald
Province, Year and Urban Pct. by Year Fixed Effects												
Log GDP	.210*	.349	.402**	1.853***	.380***	-.052	.194	.356	.412**	1.791***	.286*	-.022
	(.125)	(.225)	(.189)	(.512)	(.129)	(.199)	(.125)	(.232)	(.194)	(.486)	(.160)	(.198)
Log Electricity	.134***	.188***	.194***		.213***	.109**	.129***	.189***	.198***		.191***	.120**
	(.042)	(.034)	(.031)		(.038)	(.054)	(.041)	(.037)	(.042)		(.034)	(.053)
Log Rail Freight	.027**					.034**	.033**					.038***
	(.011)					(.015)	(.014)					(.014)
Log Loans	.224***					.219**	.187*					.200**
	(.086)					(.103)	(.100)					(.099)
Log Total Freight		-.026			-.025			-.010			-.015	
		(.033)			(.036)			(.035)			(.041)	
Log Retail Sales		.173	.072			.384		.045	-.019			.309
		(.280)	(.224)			(.255)		(.334)	(.251)			(.293)
Log VAI		-.152	-.217*					-.177	-.226*			
		(.159)	(.122)					(.155)	(.125)			
Log FAI		.174***	.170**					.186***	.174**			
		(.062)	(.067)					(.068)	(.071)			
Log Exports		.038	.038			.027		.027	.031			.024
		(.029)	(.030)			(.029)		(.041)	(.042)			(.040)
Log Imports		-.026						-.031				
		(.033)						(.033)				
Log Steel Production		.025						.019				
		(.033)						(.034)				
Log RE Investment						.025						.033
						(.026)						(.029)
Log Air Thrput						.101						.078
						(.076)						(.071)
Log LS GDP				-1.446***						-1.441***		
				(.551)						(.533)		
Log Floor Space					.020						.058	
					(.040)						(.048)	
Log Floor Starts						.005						.004
						(.016)						(.026)
Log Passengers					.012						.008	
					(.024)						(.022)	
Within R2	.20	.23	.22	.15	.17	.24	.17	.20	.19	.13	.15	.19
Prediction	7.2	4.7	5.6	9.8	6.9	6.8	6.9	5.0	5.9	9.7	6.5	6.4
CI For Prediction	(6.7,9.2)	(4.1,8.3)	(4.4,8.5)	(9.2,10.4)	(6.1,8.4)	(5.7,10.3)	(6.5,9.1)	(3.9,10.5)	(4.6,9.8)	(8.9,10.4)	(5.9,7.8)	(5.7,10.6)
Number of Obs.	270	270	270	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27

Table V presents regressions of log light intensity on log provincial GDP and components of the indices in the column headings. Standard errors clustered on province in parentheses. All predictions computed using the method explained in Section 4. All independent variables are from the CEIC and listed in the note to Table I; see the Data Appendix for details.

(V)

Table VI

China: Alternative Indices and Li Keqiang Components												
	No Urban Pct. by Year FE						*Urban Pct. by Year FE*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Base Line	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS1	Pseudo CAP	Pseudo Fernald	Base Line	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS1	Pseudo CAP	Pseudo Fernald
	Province and Year Fixed Effects											
Log GDP	.210*	.164	.215	1.205***	.182	-.052	.194	.187	.238	1.246***	.089	-.022
	(.125)	(.216)	(.190)	(.462)	(.126)	(.199)	(.125)	(.221)	(.196)	(.468)	(.154)	(.198)
Log Electricity	.134***	.118**	.125***	.099**	.119**	.109**	.129***	.126***	.137***	.100**	.098**	.120**
	(.042)	(.047)	(.040)	(.043)	(.046)	(.054)	(.041)	(.042)	(.044)	(.041)	(.042)	(.053)
Log Rail Freight	.027**	.030	.029	.023**	.036***	.034**	.033**	.031*	.031*	.027**	.048***	.038***
	(.011)	(.020)	(.018)	(.010)	(.010)	(.015)	(.014)	(.018)	(.017)	(.012)	(.015)	(.014)
Log Loans	.224***	.206**	.208**	.235**	.279***	.219**	.187*	.180*	.181*	.199*	.240**	.200**
	(.086)	(.092)	(.089)	(.095)	(.096)	(.103)	(.100)	(.097)	(.096)	(.108)	(.108)	(.099)
Log Total Freight		-.027			-.034			-.014			-.023	
		(.035)			(.033)			(.033)			(.037)	
Log Retail Sales		.269	.159			.384		.156	.076			.309
		(.294)	(.251)			(.255)		(.322)	(.257)			(.293)
Log VAI		-.123	-.186					-.144	-.195			
		(.150)	(.119)					(.145)	(.121)			
Log FAI		.165**	.165**					.178***	.171**			
		(.064)	(.067)					(.068)	(.071)			
Log Exports		.030	.029			.027		.023	.026			.024
		(.031)	(.032)			(.029)		(.040)	(.041)			(.040)
Log Imports		-.019						-.025				
		(.030)						(.029)				
Log Steel Production		.029						.024				
		(.032)						(.032)				
Log LS GDP				-1.063**						-1.121**		
				(.478)						(.499)		
Log Floor Space					.056						.097**	
					(.034)						(.042)	
Log Passengers					.031						.023	
					(.028)						(.027)	
Log RE Investment						.025						.033
						(.026)						(.029)
Log Air Thrput						.101						.078
						(.076)						(.071)
Log Floor Starts						.005						.004
						(.016)						(.026)
Within R2	.20	.26	.26	.22	.22	.24	.17	.23	.22	.20	.20	.19
P-value 3 coeffs equal	0	0	0	0	0	.01	.01	0	0	.03	.12	.02
Prediction	7.2	5.6	5.8	10.1	7.3	6.8	6.9	5.7	5.6	10.0	6.9	6.4
CI For Prediction	(6.7,9.2)	(4.7,9.7)	(5.1,10.1)	(7.4,10.2)	(6.0,10.0)	(5.8,10.0)	(6.5,9.1)	(4.3,10.6)	(4.9,10.3)	(7.5,10.2)	(5.6,10.2)	(5.8,10.5)
Number of Obs.	270	270	270	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27

(VI)

Table VI presents regressions of log light intensity on log provincial GDP, the components of the Li Keqiang index and components of the indices in the column headings. Standard errors clustered on province in parentheses. All predictions computed using the method explained in Section 4. All independent variables are from the CEIC and listed in the note to Table I; see the Data Appendix for details.

8.2 Results with Deflators as Covariates

Table VII

	China: Alternative Indices and Deflators											
	No Urban Pct. by Year FE						*Urban Pct. by Year FE*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald
	Province and Year Fixed Effects											
Log Electricity	.138***	.145***	.178***		.253***	.049	.133***	.158***	.199***		.217***	.072
	(.044)	(.045)	(.046)		(.046)	(.053)	(.043)	(.049)	(.052)		(.048)	(.051)
Log Rail Freight	.031**					.031**	.035**					.035***
	(.012)					(.013)	(.015)					(.011)
Log Loans	.262***					.236**	.221**					.225*
	(.099)					(.118)	(.111)					(.115)
Log Total Freight		-.024			.022			-.003			.020	
		(.037)			(.036)			(.036)			(.041)	
Log Retail Sales		.562***	.481***			.552***		.574**	.540**			.642***
		(.196)	(.171)			(.197)		(.259)	(.215)			(.226)
Log VAI		.012	-.040					-.012	-.044			
		(.141)	(.120)					(.137)	(.113)			
Log FAI		.128*	.122					.140**	.127*			
		(.069)	(.076)					(.071)	(.077)			
Log Exports		.028	.026			.025		.028	.030			.033
		(.026)	(.030)			(.026)		(.037)	(.038)			(.033)
Log Imports		-.048						-.050*				
		(.029)						(.029)				
Log Steel Production		.045						.039				
		(.032)						(.036)				
Log LS GDP				.482***						.402**		
				(.135)						(.170)		
Log Floor Space					.060						.092**	
					(.045)						(.042)	
Log Passengers					-.002						-.002	
					(.024)						(.022)	
Log RE Investment						.002						.007
						(.025)						(.030)
Log Air Thrput						.046						.025
						(.063)						(.057)
Log Floor Starts						-.007						-.012
						(.020)						(.029)
Log FAI Price Index	.657	.177	.055			-.045	.659	.202	.084			.007
	(.400)	(.498)	(.466)			(.365)	(.404)	(.494)	(.476)			(.373)
Log Retail Price Index		1.039*	1.011			1.269***		1.137*	1.104*			1.445***
		(.609)	(.618)			(.485)		(.614)	(.620)			(.544)
Within R2	.21	.24	.23	.07	.14	.26	.18	.22	.20	.05	.13	.23
Prediction	7.7	8.8	8.8	6.8	7.1	9.4	7.7	9.0	8.8	6.8	6.8	9.4
CI For Prediction	(6.5,9.5)	(6.0,9.8)	(6.1,9.7)	(6.8,6.8)	(6.4,8.4)	(7.9,10.4)	(6.4,9.3)	(6.2,10.2)	(6.6,9.7)	(6.8,6.8)	(6.4,7.9)	(8.2,10.2)
Number of Obs.	270	270	270	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27

(VII)

Table VII presents regressions of log light intensity on components of the indices in the column headings. Standard errors clustered on province in parentheses. All predictions computed using the method explained in Section 4. All independent variables are from the CEIC and listed in the note to Table I (deflators described in the text); see the Data Appendix for details.

Table VIII

China: Alternative Indices, GDP and Deflators												
	No Urban Pct. by Year FE						*Urban Pct. by Year FE*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald	Li Keqiang	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald
	Province and Year Fixed Effects											
Log Real GDP	.213	.795**	.832**	2.017***	.402***	.120	.212	1.077***	1.025**	2.011***	.321*	.123
	(.187)	(.353)	(.327)	(.553)	(.148)	(.226)	(.188)	(.407)	(.400)	(.591)	(.185)	(.237)
Log Electricity	.134***	.142***	.169***		.216***	.060	.132***	.136***	.172***		.195***	.079
	(.047)	(.040)	(.044)		(.038)	(.054)	(.046)	(.047)	(.050)		(.039)	(.051)
Log Rail Freight	.026**					.031**	.031**					.036***
	(.011)					(.013)	(.014)					(.012)
Log Loans	.226**					.244**	.190*					.230**
	(.091)					(.109)	(.102)					(.105)
Log Total Freight		.003			-.025			.024			-.015	
		(.038)			(.036)			(.044)			(.040)	
Log Retail Sales		.334	.185			.582***		.080	-.023			.637***
		(.266)	(.247)			(.223)		(.293)	(.278)			(.245)
Log VAI		-.272	-.314**					-.381**	-.381**			
		(.177)	(.140)					(.185)	(.158)			
Log FAI		.135**	.128*					.154**	.138**			
		(.064)	(.068)					(.062)	(.068)			
Log Exports		.040	.048*			.020		.029	.043			.026
		(.026)	(.026)			(.027)		(.036)	(.035)			(.035)
Log Imports		-.037						-.047				
		(.031)						(.032)				
Log Steel Production		.052						.056				
		(.039)						(.039)				
Log LS GDP				-1.548***						-1.573***		
				(.521)						(.527)		
Log Floor Space					.020						.058	
					(.040)						(.048)	
Log Passengers					.012						.008	
					(.024)						(.023)	
Log RE Investment						.001						.006
						(.028)						(.034)
Log Air Thrput						.043						.021
						(.075)						(.069)
Log Floor Starts						-.011						-.014
						(.018)						(.028)
Log Deflator	.084	-.143	.054	1.879***	.355	-.194	.050	-.258	.025	1.814***	.246	-.174
	(.252)	(.338)	(.265)	(.496)	(.271)	(.288)	(.256)	(.352)	(.271)	(.476)	(.277)	(.283)
Log FAI Price Index	.571	.298	.080			.024	.596	.458	.177			.072
	(.422)	(.496)	(.440)			(.364)	(.448)	(.524)	(.468)			(.393)
Log Retail Price Index		1.213**	1.121**			1.431***		1.206**	1.058*			1.568***
		(.566)	(.550)			(.494)		(.542)	(.550)			(.549)
Within R2	.21	.27	.26	.15	.17	.27	.18	.26	.23	.13	.15	.24
Prediction	6.8	7.8	7.8	9.6	6.6	9.4	6.9	7.4	7.6	9.7	6.4	9.3
CI For Prediction	(5.9,8.9)	(4.6,9.4)	(6.4,9.4)	(8.7,10.0)	(6.2,7.9)	(8.1,10.4)	(5.8,9.0)	(5.1,9.2)	(6.4,9.3)	(8.7,10.1)	(6.1,7.9)	(8.1,10.2)
Number of Obs.	270	270	270	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27

Table VIII presents regressions of log light intensity on log provincial GDP and components of the indices in the column headings. Standard errors clustered on province in parentheses. All predictions computed using the method explained in Section 4. All independent variables are from the CEIC and listed in the note to Table I (deflators described in the text); see the Data Appendix for details.

(VIII)

Table IX

China: Alternative Indices, Li Keqiang Components and Deflators	*No Urban Pct. by Year FE*							*Urban Pct. by Year FE*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Base Line	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald	All Controls	Base Line	Pseudo Barclays	Pseudo Bloomberg	Pseudo LS	Pseudo CAP	Pseudo Fernald	All Controls
Province and Year Fixed Effects														
Log Real GDP	.213	.570*	.590**	2.868**	.209	.120	.767	.212	.832**	.750**	3.015**	.144	.123	1.057
	(.187)	(.316)	(.290)	(1.275)	(.161)	(.226)	(1.806)	(.188)	(.372)	(.382)	(1.341)	(.182)	(.237)	(1.767)
Log Electricity	.134***	.063	.088*	.089*	.122**	.060	.061	.132***	.065	.099**	.089*	.106**	.079	.065
	(.047)	(.054)	(.048)	(.050)	(.051)	(.054)	(.042)	(.046)	(.050)	(.049)	(.048)	(.047)	(.051)	(.041)
Log Rail Freight	.026**	.022	.025*	.019**	.033***	.031**	.016	.031**	.020	.027**	.023**	.045***	.036***	.019
	(.011)	(.014)	(.013)	(.008)	(.012)	(.013)	(.017)	(.014)	(.012)	(.013)	(.009)	(.016)	(.012)	(.018)
Log Loans	.226**	.229**	.227**	.242***	.281***	.244**	.258***	.190*	.210**	.205**	.205*	.245**	.230**	.239**
	(.091)	(.106)	(.103)	(.093)	(.097)	(.109)	(.098)	(.102)	(.100)	(.102)	(.105)	(.106)	(.105)	(.096)
Log Total Freight		.001			-.035		-.016		.020			-.023		.001
		(.038)			(.036)		(.042)		(.040)			(.039)		(.043)
Log Retail Sales		.448	.307			.582***	.443		.253	.163			.637***	.368
		(.287)	(.277)			(.223)	(.291)		(.292)	(.292)			(.245)	(.307)
Log VAI		-.230	-.266**				-.244		-.325*	-.318**				-.282*
		(.162)	(.131)				(.156)		(.172)	(.155)				(.171)
Log FAI		.121*	.117*				.163**		.140**	.127*				.165**
		(.063)	(.065)				(.068)		(.061)	(.065)				(.067)
Log Exports		.031	.038			.020	.028		.027	.039			.026	.030
		(.027)	(.027)			(.027)	(.028)		(.033)	(.033)			(.035)	(.034)
Log Imports		-.033					-.026		-.043					-.035
		(.029)					(.035)		(.029)					(.035)
Log Steel Production		.054					.058		.059					.062
		(.039)					(.039)		(.038)					(.040)
Log LS GDP				-2.524**			-.202				-2.674**			-.364
				(1.243)			(1.691)				(1.303)			(1.726)
Log Floor Space					.045		.021					.087**		.037
					(.035)		(.053)					(.042)		(.044)
Log Passengers					.035		.022					.028		.020
					(.027)		(.029)					(.026)		(.029)
Log RE Investment						.001	-.028						.006	-.032
						(.028)	(.029)						(.034)	(.029)
Log Air Thrput						.043	.056						.021	.028
						(.075)	(.073)						(.069)	(.069)
Log Floor Starts						-.011	-.019						-.014	-.011
						(.018)	(.023)						(.028)	(.031)
Log Deflator	.084	-.304	-.117	2.457**	.067	-.194	-.121	.050	-.398	-.131	2.578**	-.038	-.174	-.024
	(.252)	(.338)	(.271)	(1.155)	(.269)	(.288)	(1.668)	(.256)	(.335)	(.264)	(1.212)	(.278)	(.283)	(1.701)
Log FAI Price Index	.571	.251	.023	-1.027	.540	.024	.148	.596	.397	.096	-1.096	.475	.072	.153
	(.422)	(.426)	(.371)	(.895)	(.431)	(.364)	(1.051)	(.448)	(.459)	(.397)	(.920)	(.405)	(.393)	(1.088)
Log Retail Price Index		1.317***	1.245***			1.431***	1.231*		1.357***	1.236***			1.568***	1.289*
		(.459)	(.441)			(.494)	(.646)		(.463)	(.463)			(.549)	(.675)
Within R2	.21	.30	.29	.23	.23	.27	.32	.18	.29	.27	.21	.21	.24	.30
P-value 3 coeffs equal	0	.01	0	0	0	.03	.01	.01	.04	.02	.04	.08	.06	.05
Prediction	6.8	8.2	8.0	10.1	6.7	9.4	8.6	6.9	7.8	7.6	10.1	6.4	9.3	8.5
CI For Prediction	(5.9,8.9)	(5.2,10.0)	(5.8,9.5)	(8.5,10.4)	(5.6,9.2)	(8.1,10.4)	(7.0,10.6)	(5.8,9.0)	(5.5,9.9)	(5.8,9.4)	(8.5,10.3)	(5.6,9.8)	(8.1,10.2)	(6.5,10.2)
Number of Obs.	270	270	270	270	270	270	270	270	270	270	270	270	270	270
Number of Clusters	27	27	27	27	27	27	27	27	27	27	27	27	27	27

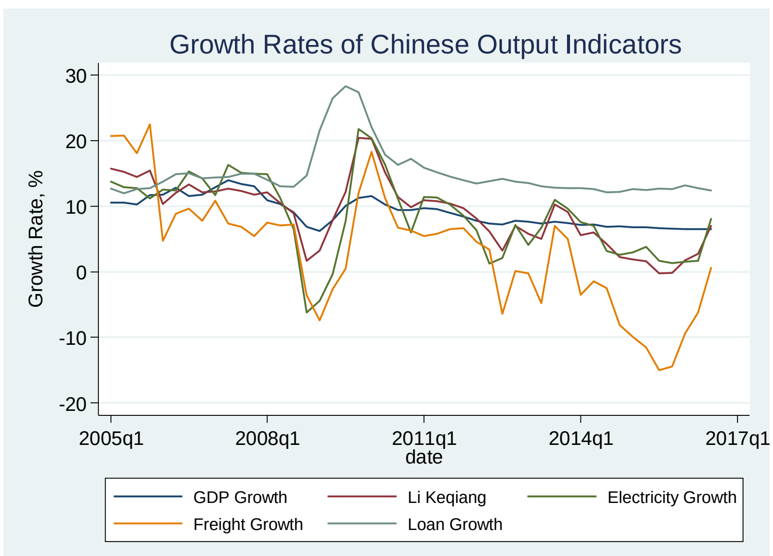
(IX)

Table IX presents regressions of log light intensity on log provincial GDP, the components of the Li Keqiang index and components of the indices in the column headings. Standard errors clustered on province in parentheses. All predictions computed using the method explained in Section 4. All independent variables are from the CEIC and listed in the note to Table I (deflators described in the text); see the Data Appendix for details.

9 Figures

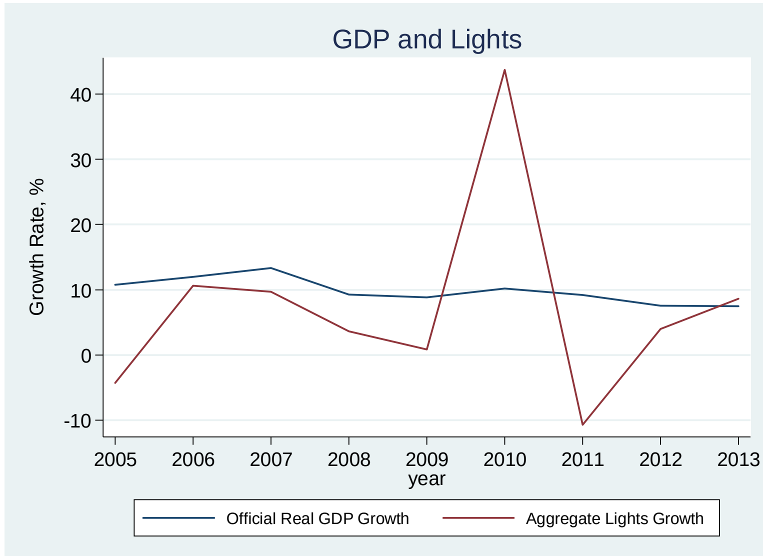
Figure 1 (1)

Figure 2 (2)



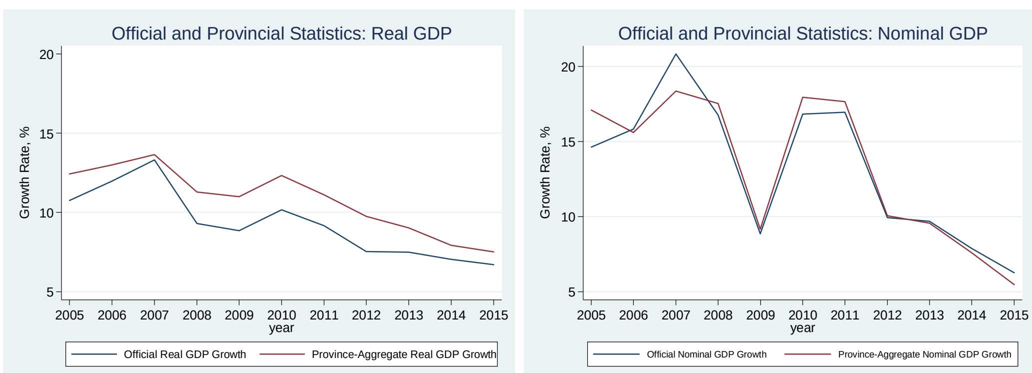
All data from CEIC

Figure 3 (3)



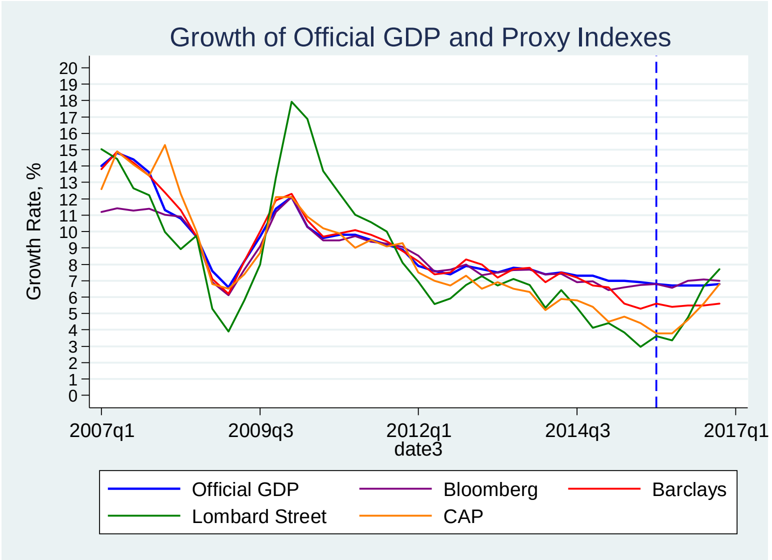
GDP data from CEIC; lights data from NOAA

Figure 4 (4)



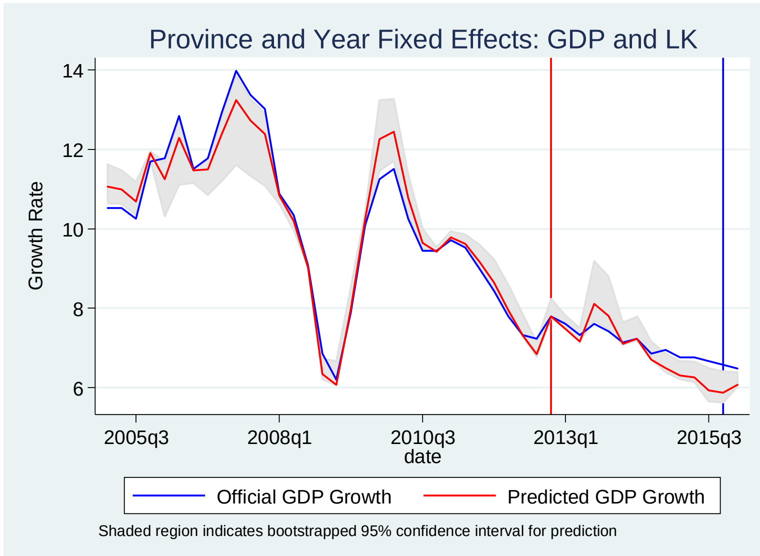
All data from CEIC

Figure 5 (5)



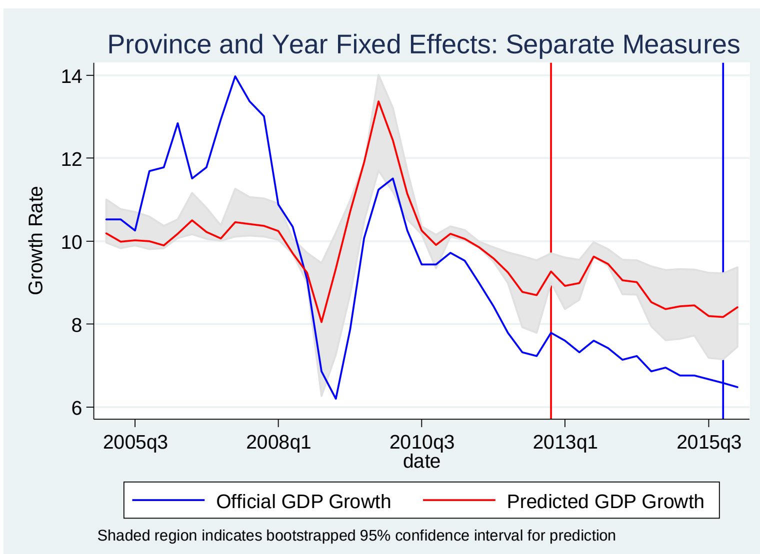
Sources: Barclays, Bloomberg, Capital Economics and Lombard Street Research

Figure 7 (7)



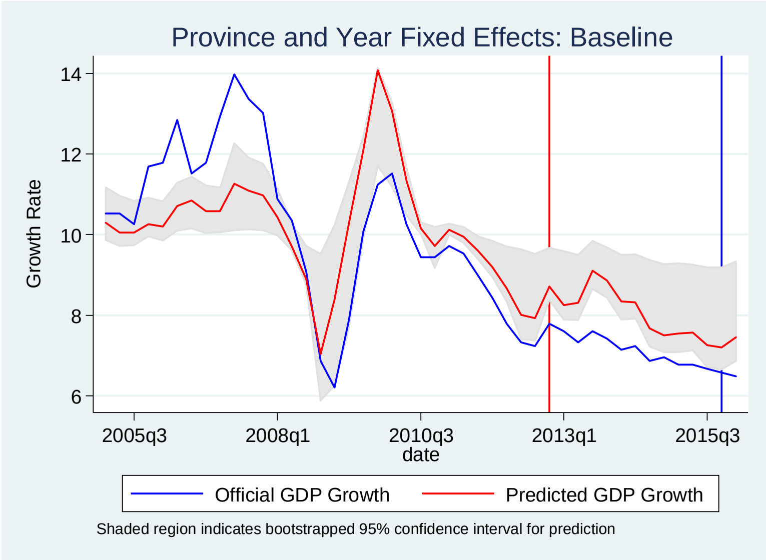
Red vertical line indicates end of calibration period; blue vertical line indicates 2015 Q4

Figure 8 (8)



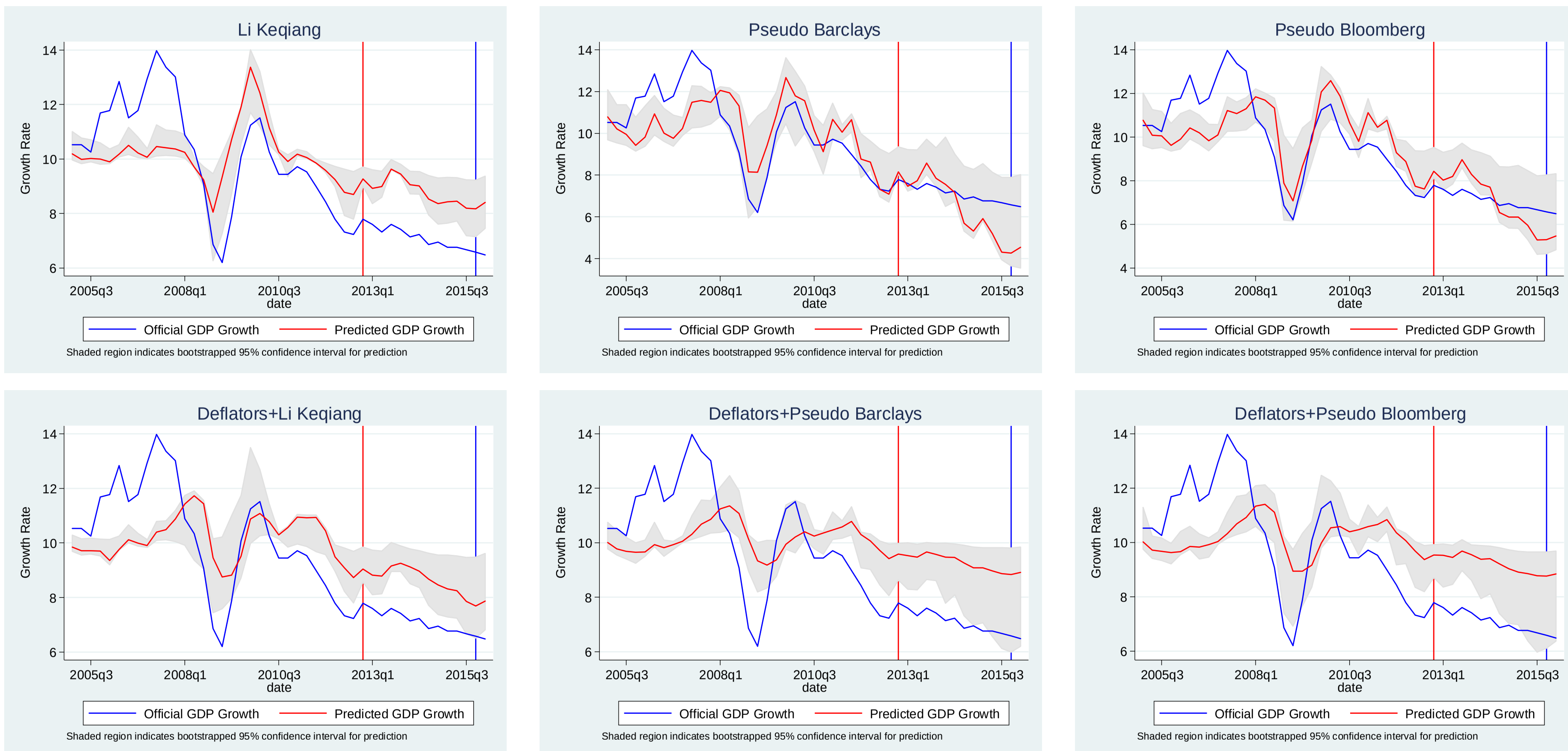
Red vertical line indicates end of calibration period; blue vertical line indicates 2015 Q4

Figure 9 (9)



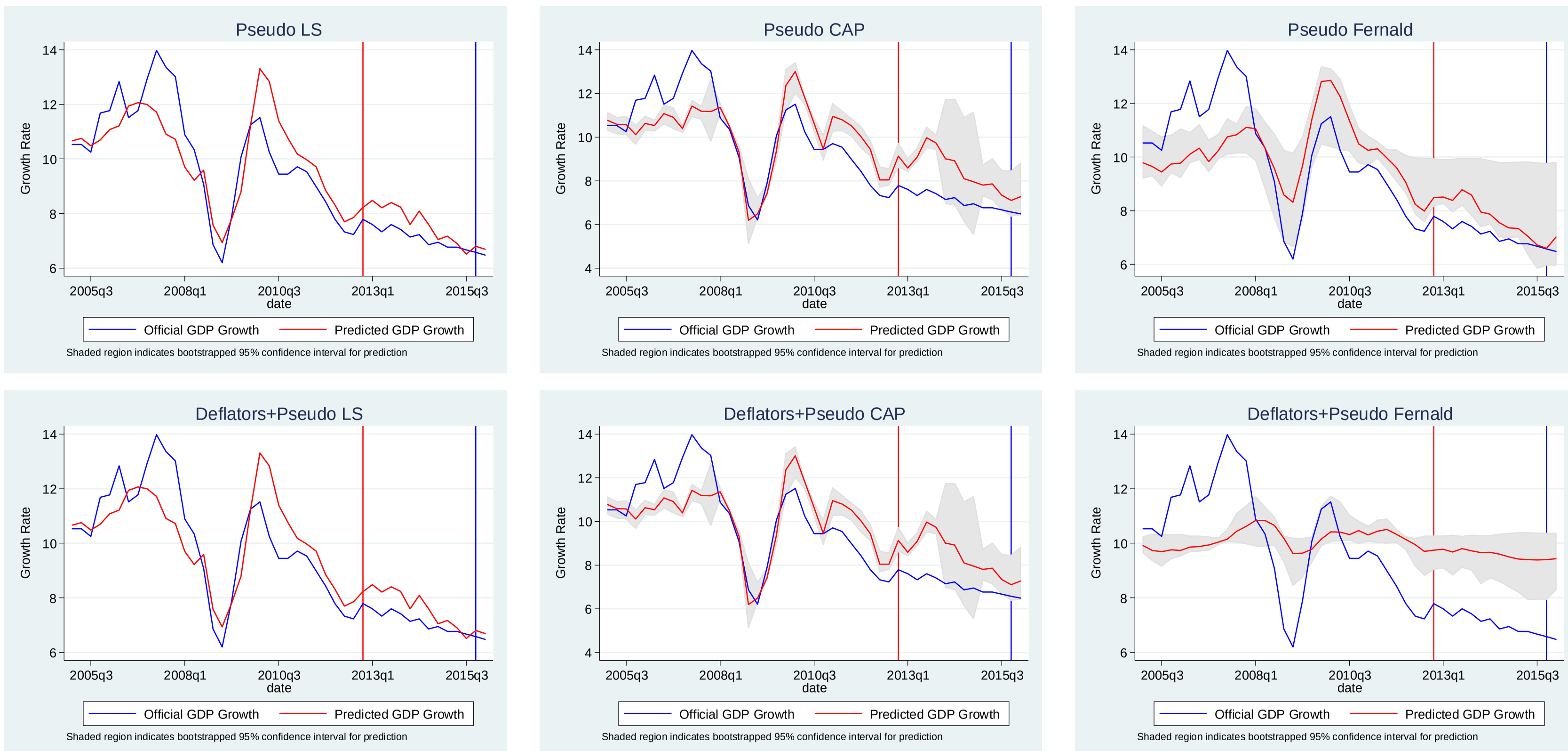
Red vertical line indicates end of calibration period; blue vertical line indicates 2015 Q4

Figure 10 (10)



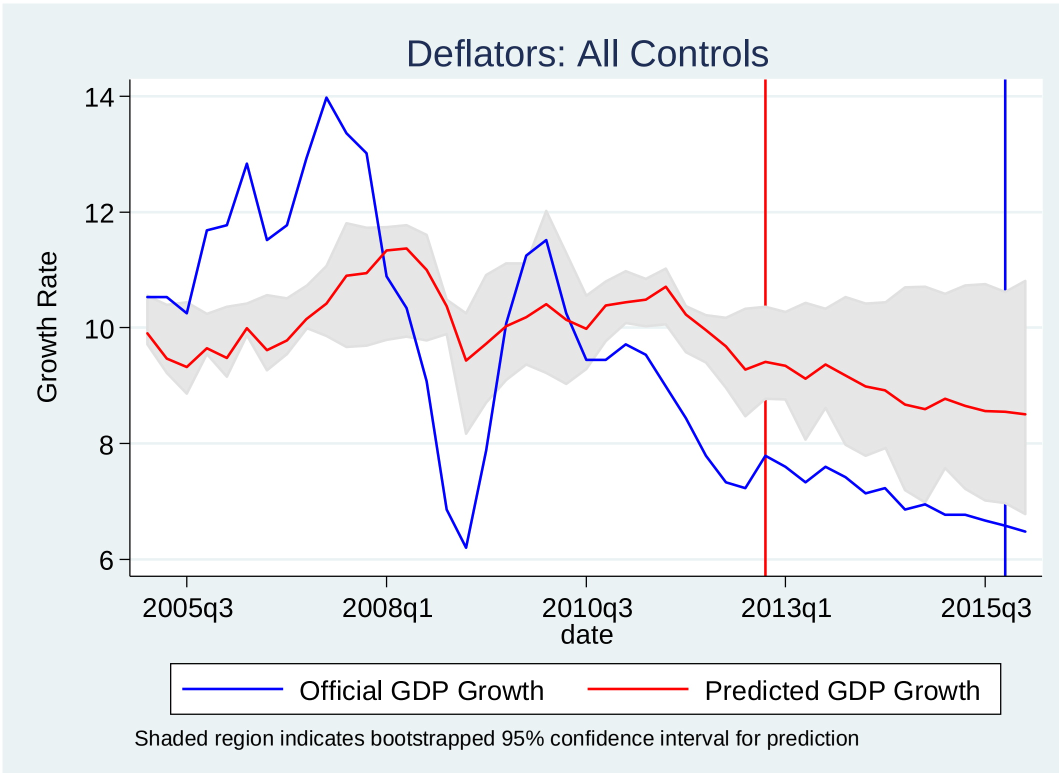
Red vertical line indicates end of calibration period; blue vertical line indicates 2015 Q4

Figure 11 (11)



Red vertical line indicates end of calibration period; blue vertical line indicates 2015 Q4

Figure 12 (12)



Red vertical line indicates end of calibration period; blue vertical line indicates 2015 Q4

10 Data Appendix

The raw data for this analysis was collected from the CEIC China Premium database, with all series, calculations, and adjustments (as relevant) described below (CEIC series codes are available on request). The main regressions using provincial and nighttime light data were conducted with annual series. National level GDP forecasts utilized quarterly series. All growth rates were calculated as year-over-year log differences.

Log GDP: Log nominal or real GDP (depending on context).

Log Primary, Log Secondary, Log Tertiary: Log nominal or real GDP (depending on context) by sector.

Log LS GDP: Log real "Pseudo Lombard St. GDP". LS GDP was calculated from official nominal GDP growth and a deflator comprised of the weighted average of the CPI, fixed asset price index, and export and import unit value indexes, using 2010 GDP weights of consumption, investment, exports, and imports. The resulting series has a correlation of 0.97 with the actual Lombard Street Research proprietary real GDP index, which was kindly provided to the authors for comparison.

Log Electricity: Log electricity production, kWh bn, at national and provincial level. Due to missing data in 1/2016 and 2/2016, data for Q1 2016 is filled-in by linear interpolation.

Log Rail Freight: Railway freight traffic, million tons.

Log Loans: Log bank loans in local currency. Provincial annual data is the annual average of monthly data. In order to help control for certain bad loan write-offs and series breaks, national level monthly data is computed from the stock of RMB loans at end 2001 and the cumulative "new increased loans in local currency" from the official aggregate financing ("total social financing") data. The national level quarterly data is computed as the three month average of this monthly data.

Log Total Freight: Log total freight traffic, million tons, by air, inland water, rail, and road.

Log Retail Sales: Log retail sales of consumer goods, RMB billions, at provincial level. To mitigate problems caused by missing data in January and February, substantial discrepancies between officially published year-to-date and monthly data in levels, and substantial discrepancies between officially published levels and percent changes, the following methodology is used to construct the national quarterly series: A nominal index in levels is constructed from the officially published seasonally adjusted monthly change from 1/2011 to the present, and backwards from 12/2010 to 1/2004 using the officially published percent year-over-year change series, which is regarded as being adjusted to make data comparable to earlier readings. It should be noted that the methodology underlying the officially published percent change is non-transparent, appears to have undergone numerous series breaks since the mid-1990s, and in years prior to 2011 at least may have included some adjustments for price increases. In practice, during the period 2005 to 2010 the officially published year-over-year percent change is always nearly identical to the percent change computed from officially published levels except for 2005 and 2010, when the officially published percent change is substantially lower and smoother. The quarterly index is computed as the three-month average of this monthly index.

Log VAI: Log real value added of industry. For provincial annual data indexes in levels are computed from official real percent growth data. To minimize missing data, for national level data an index is levels is computed from official monthly seasonally adjusted changes from 1/2011 to the present, and backwards from 12/2010 to 1/2004 from the officially published real percent year-over-year changes. The quarterly index is computed as the three-month average of the monthly data. Due to missing data in January, the first quarters of 2004 to 2011 includes only February and March.

Log FAI: Log of nominal fixed asset investment, RMB millions. Quarterly national level series is computed from the officially reported year-to-date data.

Log Exports: Log exports, by location of exporter, USD millions, at national and provincial level. Converted to RMB at the average annual or quarterly exchange rate, as relevant.

Log Imports: Log imports, by location of importer, USD millions, at national and provincial level. Converted to RMB at the average annual or quarterly exchange rate, as relevant.

Log Steel Production: Log industrial production of steel products, thousand tons, at provincial level from China Iron and Steel Association, and at national level from National Bureau of Statistics. The quarterly national level series is computed from the officially reported year-to-date data.

Log Floor Space: Log of floor space under construction, commodity building residential and non-residential, square meters, at national and provincial level. Annual data is floor space under construction as of December. Quarterly is floor space under construction as of the end of each quarter.

Log Floor Start: Log of floor space started, commodity building residential and non-residential, square meters. Annual series is cumulative starts through December of each year. Quarterly series is floor space started in each quarter computed from officially published year-to-date monthly data.

Log Passengers: Log of transport passenger traffic by air, road, water, and rail, millions of persons.

Log RE Investment: Log of real estate investment, RMB millions. Quarterly series is computed from officially published year-to-date data.

Log Air Thrput: Log of airport passenger traffic. Provincial data is "airport passenger throughput" in thousands of persons and national data is "air passenger traffic" in millions of persons.

Log Deflator: Implicit overall GDP deflators computed from nominal and real GDP series for both annual provincial GDP and quarterly national GDP.

Log FAI Price Index: Log of the fixed asset price index at the national level.

Log Retail Price Index: Log of the retail price index at the national level.