Validating China's output data using satellite observations*

Stephen D. Morris†
Department of Economics
Bowdoin College

Junjie Zhang‡
School of Global Policy and Strategy
UC San Diego

Qiang Zhang§
Center for Earth System Science
Tsinghua University

April 20, 2016

Abstract

Can officially reported output statistics be externally validated using other verifiable signals of economic growth? Satellite measured readings of nitrogen oxide emissions – a byproduct of combustion – are forwarded for this purpose. Then, a statistical methodology is proposed which allows one to utilize this and other signals, including nighttime luminosity, towards the construction of confidence intervals for output. Finally, the problem of validating China's reported GDP during two recent economic downturns is considered. Reported figures during the Great Recession period are validated for some subnational regions, but not others, including Chongqing under Bo Xilai's now known to be unscrupulous leadership.

*The Online Appendix to this paper is available at `https://goo.gl/uEkDti`. Datasets and replication materials are freely available at `https://goo.gl/aUEs4J`. The authors acknowledge funding from the Policy Design and Evaluation Lab at the School of Global Policy and Strategy, UC San Diego and thank seminar participants at Bates College, Bentley University, Bowdoin College, Middlebury College, UC San Diego, and the 2014 Workshop in Macroeconomic Research at Liberal Arts Colleges at Wesleyan University.

†Bowdoin College, Department of Economics, 9700 College Station, Brunswick, ME 04011. `smorris@bowdoin.edu`.

‡University of California, San Diego, School of Global Policy and Strategy, 9500 Gilman Drive, MC 0519, La Jolla, CA 92093-0519. `junjiezhang@ucsd.edu`.

§Tsinghua University, Center for Earth System Science, Beijing, China, 100084. `qiangzhang@tsinghua.edu.cn`.

1 Introduction

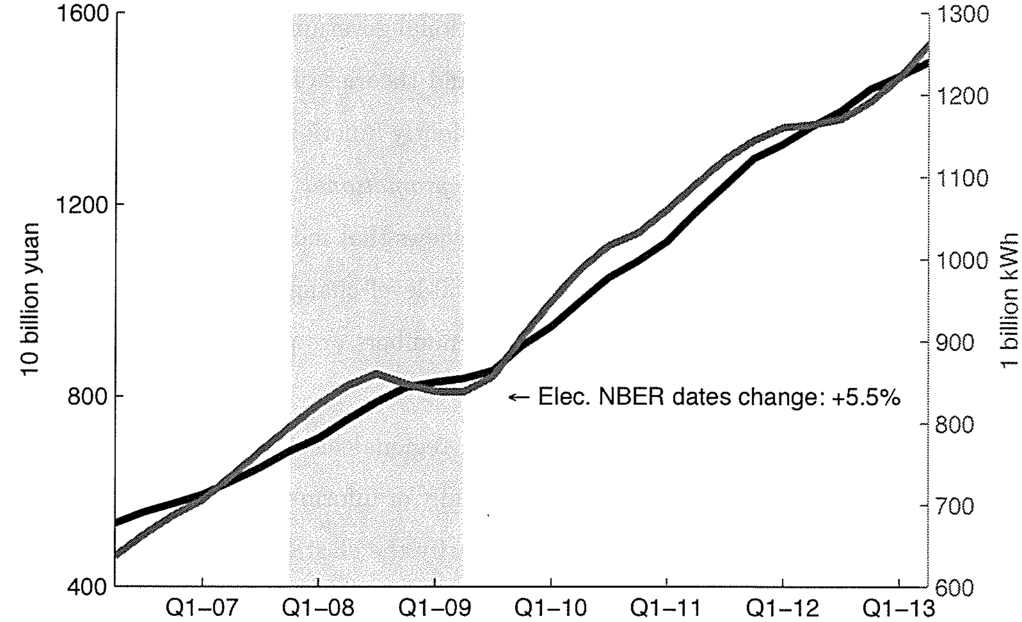
Can officially reported output statistics be externally validated using other verifiable signals of economic growth? This question is especially pertinent to Chinese series, which have long been cast under suspicion by economists. While opaque statistical methods employed in data collection represent one cause for concern,[1] another more insidious issue is systematic data manipulation. Such man-made anomalies have historically been pronounced during recessionary episodes in particular, when they become relatively more politically opportune (Wallace, 2014). For example, during the height of the Asian Financial Crisis in 1998, China reported a relatively robust 7.8% year-on-year rate of growth. This figure appears to be a clear aberration not only regionally, but also in the context of sharper contemporaneous declines in energy consumption statistics (Rawski, 2001). Similar concerns arose during the Global Financial Crisis and ensuing Great Recession period of 2007-2009, an era during which output data appears ex-post to have been too smooth (Nakamura et al., 2016). But pairwise incompatibilities with other conventional indicators of growth, such as electricity generation, were evidently more muted during this latter episode. Indeed, both indicate a healthy rise in economic activity over the NBER recession dates, 2007 Q4 - 2009 Q2 (Figure 1 (a)). Perhaps due to this redundancy of information, economists have in this case come to more mixed conclusions as to the viability of reported statistics; Fernald et al. (2013) argue that industrial production indices corroborate reported GDP data, while the opposite case had previously been made by Koech and Wang (2012). Given that the Chinese economy now seems poised for another period of protracted sluggishness, the question of reliably validating officially reported output statistics plainly deserves further attention.

Statistics which are merely indicative of aggregate output are usually called *signals* of economic growth. Fundamentally, their relationship with GDP is not one-to-one. In order to make substantive conclusions regarding output from any signal, it must first be transformed into a *proxy*. To make this transformation, one must know or otherwise be able to estimate the *elasticity* of growth with respect to the signal. The most notable proxy for economic growth in China is the "Li Index," named after Chinese Premier Li Keqiang. As then Party Secretary of Liaoning province, Li was widely cited through Wikileaks cables as using electricity consumption, volume

[1] Chang et al. (2015) have only recently developed an annual and quarterly macroeconomic time series dataset for China comparable with those commonly utilized in empirical research on Western economies.

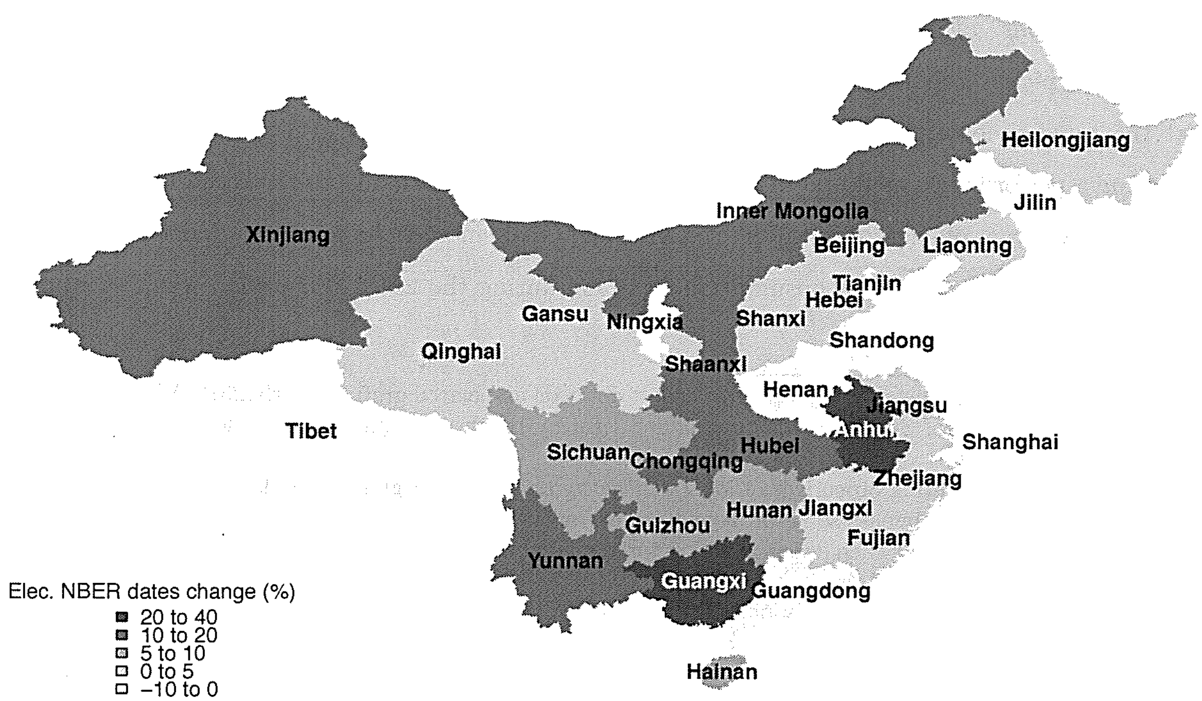
Figure 1: Reported GDP and electricity generation: 2006-2013.

(a) Reported GDP vs. electricity generation: All China.



Notes: Quarterly four-quarter rolling average, 2006Q1 - 2013Q3. GDP: Left axis, black. Electricity: Right axis, red. Shaded: NBER recession dates (07Q4 - 09Q2). GDP NBER dates change: +22.3%.

(b) Electricity generation change over NBER dates: By region.



of rail cargo, and the amount of loans disbursed as signals from which to infer his measure (e.g. Batson, 2010). The fact that Party elite would be forced to rely on such a proxy offers perspective on a key aspect of Chinese political economy. The national government of China naturally desires accurate statistics for the purpose of policymaking, and theories of manipulation at this vantage are usually refuted (Chow, 2006).[2] However, independently functioning provincial and municipality officials, vying for power, involve themselves in a *promotional tournament* for advancement; their ascent is contingent upon annual performance reviews that include GDP statistics from their district (Chen et al., 2005).[3] Inasmuch, it is the local-level change in signals which is relevant for proxy construction, as this is the scope at which numbers are potentially managed; variance across regions may be substantial (Figure 1 (b)).[4]

And yet, is there potentially something awry in regional electricity generation data itself? While we have no way of knowing exactly how Li formally or informally transformed these regional signals, two assumptions are always necessary to identify the elasticity required for inference: (1) The signal is reliable, and (2) the relationship with output is known. With respect to industrial production indices, a case may be made for the latter qualification. But given the limelight shed on the Li index, the former can no longer be assumed. Indeed, a common viewpoint is that data manipulation has become more cunning in recent years (Koch-Weser, 2013).[5] Willful "reconciliations" of reported output and conventional signals like electricity generation – which are also typically reported by individual regions – are perhaps too speculative to assume, but at the same time can not simply be assumed away. Moreover, energy consumption itself has become a performance measure for promotion of local officials, thanks to the recently introduced mandatory target of energy intensity (Sinton, 2002; Ghanem and Zhang, 2014). Indeed, Liu et al. (2015) estimate that energy consumption in China during the 2010-2012 period was 10% above officially reported statistics.[6]

[2]The national government has publicly ventured to cut down on associated graft. In early 2016, Wang Baoan, director of China's National Bureau of Statistics, was put under scrutiny by the Communist Party for what it called "serious violations."

[3]As evidence of the competing objectives of local and national officials, beginning in 1998, China's National Bureau of Statistics publicly began to bypass some provincial governments in data accumulation.

[4]Analogues to Figure 1 (a) for each subnational region depicted in Figure 1 (b) are given in the Online Appendix to the this paper, Appendix A, Figure A.1.

[5]Holz (2014) has argued that previously offered evidence of data manipulation in China is woefully not compelling, and that China's National Bureau of Statistics has the freedom to doctor figures in a "virtually undetectable" manner.

[6]Another potential cause of apparent data inconsistencies is that the structural relationship between electricity

Given these concerns, it is desirable to creatively seek out signals which are guaranteed to be free of systematic distortion. Fernald et al. (2015), for example, ingeniously use the volume of imports *to* China, as reported by trading partners.[7] The benchmark in the realm of autonomous growth measurement more generally is nighttime luminosity, or "night lights." This signal is not only naturally indicative of energy consumption, but measured by orbiting international satellite instruments, and publicly available.[8] Henderson et al. (2012) pioneered the application of this dataset to formally producing *combined measures* of economic growth – optimal weighted averages of proxy and reported growth – in areas where data collection is otherwise challenging. Unfortunately, a follow-up study by Chen and Nordhaus (2011) casts doubt on the usefulness of luminosity towards constructing proxies for growth in China or other middle income countries. Their conclusion is that luminosity is useful mainly in countries with little or no operative systems for data collection. In addition, luminosity data series are available at no higher frequency than an annual basis, making them unsuitable for the construction of combined measures at the more commonly studied quarterly interval. Furthermore, the DMSP-OLS luminosity dataset currently only ranges from 1992 to 2013, an annual time dimension of just $T = 22$. Inasmuch, one must generally exploit the cross sectional, rather than time series dimension of the data, for estimation. As substantiated below, this restriction may be problematic when considering a technologically diverse country like China.

Still, the proposal of utilizing remote sensing technology to retrieve useful signals for combined measure construction is promising. This paper forwards the use of nitrogen oxide (NO_X) emissions retrieved from multiple international satellite instruments. Satellite measurements of NO_X, a byproduct of anthropogenic sources primarily including combustion but also biomass burning, have already been exploited to verify Chinese energy consumption and pollution in atmospheric science (Akimoto et al., 2006; Lin and McElroy, 2011). High spatial resolution and sampling frequency enables one to assemble a longitudinal data set which like luminosity is guaranteed to be free of political influence for a large portion of the Earth's surface. But unlike luminosity, NO_X

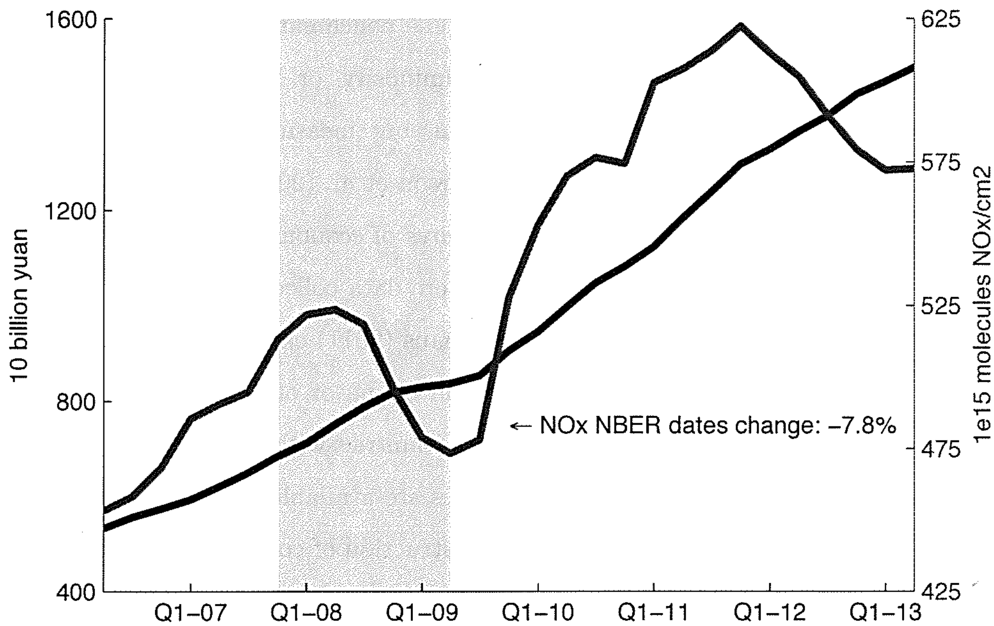
consumption and growth in China may naturally vary during downturns (Lin and Liu, 2016). However, this opposing perspective appears to be a minority viewpoint.

[7] An alternative FAVAR based methodology for dealing with the same problem is put forth in Fernald et al. (2014).

[8] The United States Air Force's Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) nighttime lights global annual panel available at 30 arc second grids between -180 180 degrees longitude and -65 to 75 latitude is publicly available at `http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`.

Figure 2: Reported GDP and NO_X emissions: 2006-2013.

(a) Reported GDP vs. NO_X emissions: All China.



Notes: Quarterly four-quarter rolling average, 2006Q1 - 2013Q3. GDP: Left axis, black. NO_x: Right axis, blue. Shaded: NBER recession dates (07Q4 - 09Q2). GDP NBER dates change: +22.3%.

(b) NO_X emissions change over NBER dates: By region.

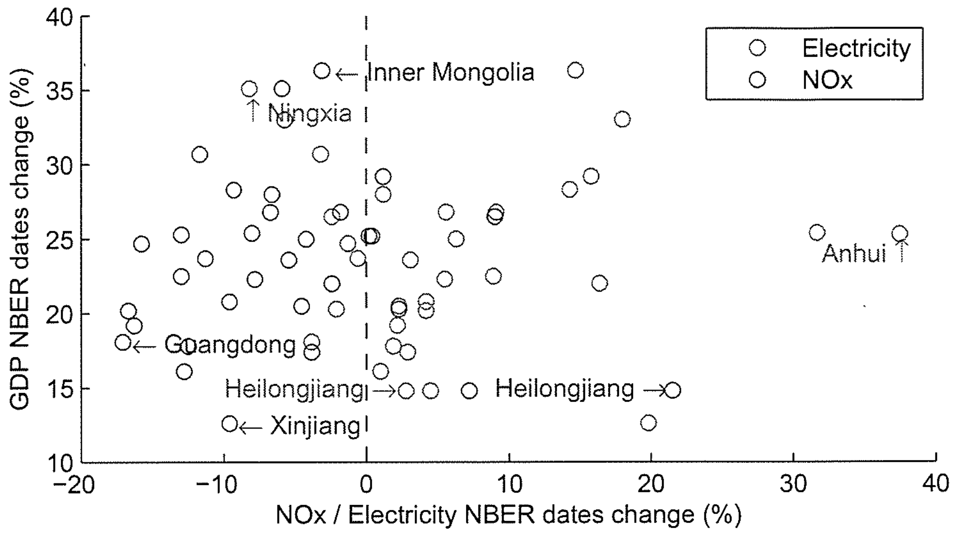
data is reliable to at least a monthly basis, allowing one to consider business-cycle fluctuations at the regional level. Furthermore, these observations range from as far back as 1996 to present, resulting in a time series dimension exploitable for estimation.[9] Figure 2 (a) replicates Figure 1 (a), but now using NO_X emissions. Recall, electricity production increased 5.5% over the NBER business cycle dates, while reported output grew 22.3%. Yet, anthropogenic emissions of NO_X *fell*, growing -7.8%. Given that NO_X emissions are produced in the generation of electricity – for example by coal-fired power plants – this observation seemingly bears some credence upon the theory that electricity generation figures may be misleading. Moreover, NO_X emissions indicate a different pattern of severity of the crisis across regions (compare Figure 2 (b) with Figure 1 (b)).[10] Specifically, NO_X emissions depict an evident geographic dependency in downturn severity which is not embodied in electricity generation figures. This further bolsters the usefulness of NO_X emissions in comparison with electricity generation insofar as this geographic dependence is to be expected (Fogli et al., 2012).

To unwind these observations more formally, the second contribution of this paper is econometric. Again, the pathbreaking luminosity study of Henderson et al. provides the starting point for this analysis. The primary assumptions utilized for estimation in that context are that the signal and true output are observed with classical measurement error, and that the elasticity of signal with respect to output is constant across regions. In other words, asymptotic results are of the large N, small T form. Yet this parsimony comes at some potentially significant practical expense to the current application. As is explained below, classical measurement error precludes the possibility of politically motivated systematic manipulation. Furthermore, the assumption that signal-output elasticities are constant across provinces and municipalities is unlikely to hold given the technological diversity of these areas. Viewing again Figure 2 (b), it is striking that NO_X emissions *increase* by a comparatively large amount in the northeastern provinces of Liaoning, Jilin, and Heilongjiang over the NBER dates. But this result is not necessarily surprising, since this greater region, Manchuria, is traditionally industrial, and coal dependent. In other words, the pass-through of emissions per dollar GDP inevitably differs regionally based solely on the primary combusted fuel source and production technology. Figure 3 underscores these regional differences

[9]The following section of the paper describes the dataset utilized herein more completely.

[10]Analogues to Figure 2 (a) for each subnational region depicted in Figure 2 (b) are given in Appendix A, Figure A.2.

Figure 3: China regional signal vs. GDP change: NBER recession dates (2007 Q4 - 2009 Q2).



in signal elasticities. It plots the percentage change in electricity generation and NO_X emissions over the NBER dates previously depicted in Figures 1 (b) and 2 (b) against respective regional percentage change in GDP. If the elasticity of either signal with respect to GDP is geographically constant across China, then one should observe an upward trend. This is dubious in either plot. Finally, Henderson et al.'s framework does not accommodate multiple signals, nor is it clear how to compute confidence intervals for the combined measure. Without such a measurement of stochastic uncertainty, it is impossible to determine when a given combined measure and officially reported output statistic are statistically distinct. Yet, that is the key question underlying the validation of any reported output statistic.

This paper develops a methodology for combined measure construction designed to address these particular concerns. First, it presents a structural model not dissimilar from the Henderson et al. framework. The distinction is that this model – which ultimately has dynamic panel data (DPD) representation – allows for systematic politically-motivated misreporting, without its assumption. Then, the paper presents large T, small N results which may be applied to the estimation of subsets of technologically similar regions. The nuisance parameters which arise in the mapping from DPD to underlying structural parameters, including the imperative signal elasticity, clarify the identifiability pitfalls researchers are prone to when they attempt to process signals in reduced form frameworks. Finally, this paper also describes how to generalize this estimator to multiple signals, and compute bootstrap confidence intervals for combined measures. These *error bands* are also immediately interpretable as confidence intervals for output. Inasmuch, they provide the analyst with a formal basis for determining which GDP statistics are statistically improbable. Finally, though the concentration of this paper is China, the empirical methodology developed herein is generally applicable to any signal or geography.

Ultimately, this new signal and methodology are applied Chinese data. Since the incentives of local officials to misreport output quantities may increase during periods of economic turmoil, the focus will be two recent trying periods: The Asian Financial Crisis of 1997-1999 and Great Recession of 2007-2009. As part of this analysis, luminosity is reconsidered, along with several indices of industrial production. The results indicate that luminosity is marginally preferred by the data to all other alternatives for combined measure construction, including NO_X, in annual data. This tends to buoy the usefulness of this signal in China, in contrast with Chen and Nordhaus's previous result. However, the primary disadvantages of luminosity are its low frequency availability and relatively short time dimension. Specifically, in order to conclude that any reported output figure is statistically improbable, one must be able to show that it falls outside of the error bands computed using a given signal. But in order to come to the strictly stronger conclusion that this discrepancy is due to political aspirations, and not just a period of remarkable growth, such an event minimally must occur at a politically opportune time: During or subsequent to a crisis period. The substantive finding is that the restrictions imposed on estimation purely by the dimensions of the luminosity panel make it difficult to utilize towards this particular ends. In part due to the need to extend the N dimension utilized in estimation with small T, and in part due to the relative smoothness of annual data, all annual signals are too uninformative to be useful towards uncovering possible episodes of manipulation.

In contrast, when considering quarterly data, NO_X emissions are however useful for pinpointing such instances of possible misreporting, and these inferences are corroborated by industrial production indices. In particular, this paper considers the experience of four major cities during the Great Recession period: Beijing, Tianjin, Shanghai, and Chongqing. A purely statistical analysis of the data alone indicates that reported GDP figures for Beijing, Tianjin, and Shanghai are supported by NO_X emissions data, while reported GDP figures from Chongqing are not. In fact, this is the time period during which Chongqing was under the stewardship of Bo Xilai, a high-flying political official who gained prestige for outstanding GDP numbers, but was famously deposed from power and sentenced to life imprisonment for corruption in 2013. GDP statistics reported by Xilai's administration were contemporaneously under scrutiny by his rivals; in the words of his predecessor Wang Yang in 2009, "Some of our GDP figures sure look rosy." (Liu, 2009). The conclusion of this analysis is therefore that NO_X emissions are useful in separating likely manipulated from accurate statistics.

This paper is structured as follows: Section 2 describes the construction of NO_X series and dataset under consideration more broadly. Section 3 develops the structural model to be utilized in the remainder of the paper. Section 4 studies the identifiability of the structural parameters in such models and introduces a new estimator for combined measures, substantiating its properties by Monte Carlo experiment. Section 5 considers combined measures of output regionally in China during crisis periods. Section 6 concludes.

2 Dataset

Nitrogen oxides (NO_X) is a generic umbrella term encapsulating both nitric oxides (NO) and nitrogen dioxides (NO_2). Emissions of these particles are produced primarily during combustion, and can be particularly pronounced in areas with large quantities of automobile traffic. Globally, fuel oil is thought to contribute to roughly 50% of anthropogenic NO_X emissions. Coal-fired power plants are another major contributor. Aside from the burning of fossil fuels, NO_X is also a byproduct of biomass burning and fertilizer, which are also clearly indicative of production. But NO_X is also naturally occurring in soil and lightning, which therefore must be controlled for to deduce anthropogenic emissions, the statistic of interest. In terms of externalities, NO_X emissions can cause smog and haze, as well as acid rain.

Anthropogenic emissions of NO_X have surged in China over the past decades coinciding with increased economic growth. Traditionally, NO_X inventories are computed using what is known as the bottom-up approach. This methodology utilizes known fuel consumption and emissions factors, or, technology-dependent degrees of intensity with which fuel is converted into pollution. But limited access to such data in China in particular, where these numbers are also part of local

Table 1: Remote instruments.

Instrument	GOME	SCIAMACHY	GOME-2	OMI
Onboard satellite	ERS-2	ENVISAT	MetOP-A/B	EOS-Aura
Time sample	1996-2003	2002-2012	2007-present	2004-present
Resolution (km^2)	30×60	30×60	40×80	13×24
Global coverage	6 days	6 days	$\sim$ 1 day	1 day

Notes: Available data sources for the retrieval of NO_2 vertical column densities from space.

figureheads' annual review, has led scientists to instead pursue what is known as the top-down approach. Simply put, by this method one attempts to deduce emissions from satellite readings (Wang et al., 2012). Specifically, satellite measurements of NO_2 vertical column densities (VCDs) are used to constrain NO_X emissions estimates via a chemical transport model relating stocks (densities) to flows (emissions). Purely human-made emissions can be determined independently via the transport model using surface activity data such as land cover, wild fire and meteorological inputs to exclude natural causes.[11] The satellite VCD readings of NO_2 utilized in this paper are described by Table 1. Anthropogenic NO_X emissions are deduced from these measurements using the transport model methodology described in Zhang et al. (2007).

How accurate are these emissions readings? In fact, amongst all tropospheric species, NO_X is arguably the easiest to measure. The short atmospheric life of NO_X means that densities are closely correlated to emissions, and therefore may be used to accurately measure them independent of wind or other obfuscation. In fact, NO_2 readings have been used to indirectly infer other more difficult to measures species, such as carbon dioxide, over China (Berenzin et al., 2013). These statistics are also clearly correlated with economic growth, a fact well known to the scientific community. For example, it has been noted that the resurgence in Chinese growth following the Global Financial Crisis is clearly evident in this data (Itahashi et al., 2014). Henceforth, referrals to "NO_X" or "NO_X emissions" always mean anthropogenic NO_X emissions.

Using these readings, two separate panels are created, described by Table 2. Both panels correspond to the same set of $N = 31$ provinces, municipalities, and autonomous regions previously depicted in Figures 1 (b) and 2 (b). But the panels differ in the time T dimension. The

[11] Because anthropogenic emissions of NO_X are mainly attributed to the combustion of fossil fuels, energy consumption can also be inferred directly using a known emissions factor.

Table 2: Dataset.

$N = 31$ regions		Units	Annual 1993-2008 ($T = 16$)	Quarterly 06Q1 - 13Q3 ($T = 31$)
Reported output	GDP	yuan	✓	✓
Satellite signals	Luminosity	watts / cm^2	✓	
	NO_X emissions	molecules / cm^2	✓	✓
Reported signals	Freight traffic	tons	✓	
	Electricity generation	kilowatt-hours	✓	✓
	Cement production	tons	✓	✓
	Steel production	tons	✓*	✓**
Crisis period			*1997-9*	*2007-9*

Notes: $N = 31$ provinces, municipalities (Beijing, Tianjin, Shanghai, Chongqing) and autonomous regions (Guangxi, Inner Mongolia, Tibet, Ningxia, Xinjiang). *Tibet missing. **Tibet, Hainan, Ningxia missing.

time periods selected purposefully subsume two economically challenging periods in China, during which systematic manipulation of data at the sub-national level may be more likely. The first time period, at the annual frequency from 1993-2008, includes the Asian Financial Crisis of 1997-1999. This annualized series is studied in part to serve as a control from which to compare the efficacy of NO_X versus nighttime luminosity, which is not available at higher frequency, in combined measure construction. In addition, several reported indices of industrial production, including two elements of the Li Keqiang Index (freight traffic and electricity generation), are included. The evolution of the annualized percentage change in these series for China as a whole over this time period is depicted in Figure 4. Characteristically, all series fall during the Asian Financial Crisis period of 1997-9, and recover however moderately thereafter.

The second time period, at the quarterly frequency from 2006 Q1 - 2013 Q3, includes the Global Financial Crisis and ensuing Great Recession of 2007-2009. Freight traffic must also be dropped, as it is not available at higher than annual frequency. Lin and McElroy (2011) previously noted that NO_X fell considerably during the 2007-9 period. Figure 7 depicts a similar contemporary drop in the annualized percentage change in each series for China as a whole. Importantly, we observe negative rates of change for NO_X during the recession period, while GDP approaches but never falls below a 0% rate of change.

In Table 2, signals which are satellite-based versus reported by individual regions are differentiated between. While it seems too speculative to hypothesize that such reported signals are "reconciled" to reported GDP statistics, as discussed in the introduction, one may nonetheless only conclude that this is necessarily not the case for satellite-based measures. For the remainder of the analysis, arguments are parsed so as to consider luminosity and NO_X as the primary signals of interest, whereas all remaining indices of industrial production are useful, though to be interpreted with a dose of caution. In other words, results stemming from these signals should be interpreted as primarily auxiliary in nature.

The key to utilizing any signal for proxy construction lies in identifying the elasticity of the signal with respect to income, β. If this signal is constant across regions $n = 1, \ldots, N$ and time periods $t = 1, \ldots, T$, then the following equality will hold with error ε_{nt}:

$$\%\Delta\text{Signal}_{nt} = \beta \times \%\Delta\text{Output}_{nt} + \varepsilon_{nt} \tag{2.1}$$

Let us consider the viability of the assumption that β is indeed constant across both n and t. First, to establish that β is constant across t, one may exploit variability across n to estimate β within each period. In Figure 5, elasticities for both luminosity and NO_X are estimated for each year in the annual sample. They are statistically indistinguishable from one another. In the Online Appendix to the this paper, Appendix A, Figure A.3, the same is shown to hold for NO_X across each quarter in the quarterly sample. This suggests that β is indeed constant across the time series dimension of the data for each area, and exploiting large T results to estimate β is valid.

But is it also true that the estimated elasticity is stable across regions? Figure 6 depicts such estimates for both luminosity and NO_X in the annual sample. These estimates strongly suggests that this latter assumption is not correct; estimated luminosity elasticities vary from $\widehat{\beta} = 0.2$ for Beijing to $\widehat{\beta} = 0.9$ for Chongqing. A similar result is indicated for NO_X in the quarterly sample by Appendix Figure A.4. Intuitively, each area is economically distinct, and variety technologies and utilized inputs result in distinct elasticities of each signal with respect to output. In this sense, β is *not* constant across n, and exploiting large N results to estimate (2.1) is unlikely to result in a consistent estimator.

One may further corroborate this result by analyzing the time dimension cross-correlations for each individual area. Appendix Table A.1 lists correlations of reported GDP with each signal, and each luminosity and NO_X emissions versus each other signal, for the annual sample. Across regions, GDP is most correlated with luminosity, with NO_X emissions a close second. GDP, luminosity, and NO_X are all least correlated with steel production. But these correlations fluctuate greatly across regions; the correlation of GDP with luminosity is 0.81 for Hainan but 0 for Sichuan. Table A.2 presents the same, excluding luminosity and freight, for the quarterly sample. Similar results hold, with GDP most correlated across regions with NO_X and electricity generation, and least correlated with concrete and steel production. Thus, in both samples there is ample evidence that signal elasticities are not constant across regions in China, a result which holds for all signals. Finally, due to the apparent low correlation of steel production with GDP, and its multiple missing values, for the remainder of the paper it is omitted in favor of the remaining signals.

Elasticities with respect to satellite-based signals seem to be remarkably stable across time, suggesting large T results are useful. But on the other hand, the same can not necessarily be said across areas within China, suggesting large N results are less so. The implications of this primitive analysis of the data are twofold. First, the DMSP-OLS luminosity panel is currently restricted to a 1992-2013 time sample, whereas monthly data for NO_X is available at longer horizon and higher frequency (Table 1). In situations such as these where it is not appropriate to exploit large N results, NO_X emissions may be useful in comparison, a hypothesis yet to be tested. Second, the fact that the time series dimension of the data is to be exploited calls for a compatible statistical methodology, which does not exist. In order to develop such a methodology, we must first develop a model consistent with the type of systematic data manipulation believed to possibly exist. The next section removes itself from the specific case of China and NO_X to present such a generally applicable framework.

3 Model

Combined measures of economic growth are composite series of conventional output data and other proxies for growth. Such proxies are created from subjectively chosen signals using a statistical procedure. While these signals contain only indirect information about economic activity, that information is deemed relatively reliable. Therefore, combined measures are forwarded as a useful alternative measure of economic growth when conventional output data alone may be erroneous.

This section offers refinements in the arena of combined measure construction pioneered by Henderson et al. (2012). The purpose of this exercise will be to develop a framework for growth measurement which is applicable to areas lacking developed statistical agencies, as those authors assumed, but also other areas where output may be misreported in a purposeful way. First, a model is proposed which incorporates serially correlated unobservable states. This generalization allows for the possibility that output data is systematically manipulated, but imposes no such restriction. Then, it is demonstrated that a combined measure remains available in this augmented setting. Finally, this setup is shown to be immediately generalizable to allow for arbitrarily many signals. This allows one to consider not only the unilateral information value of individual signals, but also the joint universe of possibilities.

3.1 Structural equations

Let Y^*_{nt} be the level of latent, or otherwise unknown, unobserved, or poorly measured output for geographical area $n = 1, \ldots, N$ in time period $t = 1, \ldots, T$. Generally, n may correspond to any arbitrary unit of area, ranging from an entire country, to a mesh grid of the finest resolution. t typically means years or quarters; higher frequency output data is not usually available or studied even for sovereignties with very accurate statistical records. Henceforth, take the convention of exponent-* to mean variables which are not directly unobservable.

Say $\widetilde{y}^*_{nt} = 100 \times \Delta \ln Y^*_{nt} \approx \%\Delta Y^*_{nt}$ evolves according to an AR(1) process $\widetilde{y}^*_{nt} = \alpha_n + \rho_y \widetilde{y}^*_{nt-1} + \varepsilon^y_{nt}$ for macroeconomic shock ε^y_{nt}. α_n are fixed effects attributable to either potential output for area n, or average mismeasurement over the time sample under consideration. Then the percentage change less-means $y^*_{nt} = \widetilde{y}^*_{nt} - E(\widetilde{y}^*_{nt})$ follows

$$y^*_{nt} = \rho_y y^*_{nt-1} + \varepsilon^y_{nt} \tag{3.1}$$

Appendix B provides an example of a simple but ubiquitous dynamic stochastic general equilibrium model which corroborates this reduced form assumption. The following standard regularity conditions are also assumed: $0 < |\rho_y| < 1$, $\varepsilon^y_{nt} \sim IWN(0, \sigma^2_y)$ independent white noise for $0 < \sigma_y < \infty$, and $\{\varepsilon^y_{nt}\}$ has finite fourth moments. The shocks are therefore not serially correlated. However, they are not necessarily uncorrelated across areas, contemporaneously. This allows for areas to share identical shocks, but at the same time does not impose any such restriction.

(3.1) is unique to the model specification of this paper, versus Henderson et al.'s setup. It is included primarily for reasons pertaining to identifiability which will become evident in the following section. However, the next three equations are largely consistent. Say that that the latent or otherwise only indirectly observable value of a signal, S^*_{nt}, is related to latent output by the relation $S^*_{nt} = Y^{*\beta}_{nt}$. Thus, β is the elasticity of the signal with respect to income. Define $\widetilde{s}^*_{nt} = 100 \times \Delta \ln S^*_{nt}$ and $s^*_{nt} = \widetilde{s}^*_{nt} - E(\widetilde{s}^*_{nt})$. s^*_{nt} are as such differenced from their own fixed effects, which, for example, may arise from an attempt to "reconcile" such signals to output. Specifically, s^*_{nt} is related to y^*_{nt} by

$$s^*_{nt} = \beta y^*_{nt} \tag{3.2}$$

$\beta \neq 0$ by assumption. This requires that the signal is in a colloquial sense, relevant. The observed or otherwise scientifically recorded value of the signal is related to this latent value with error. For example, a known issue pertaining to luminosity data is that phantom readings may occur over oceans near coastal settlements. Henceforth, this paper adopts the notation that variables without-* (s_{nt}) represent the observable datum corresponding to each with-* unobservable counterpart (s^*_{nt}).

$$s_{nt} = s^*_{nt} + \varepsilon^s_{nt} \tag{3.3}$$

$\varepsilon^s_{nt} \sim IWN(0, \sigma^2)$ for $0 < \sigma < \infty$, $\{\varepsilon^s_{nt}\}$ has finite fourth moments, and $\{\varepsilon^s_{nt}\}$ is independent of $\{\varepsilon^y_{nt}\}$. In words, signal measurement error is not serially correlated, though possibly correlated across areas. Furthermore, it is uncorrelated from macroeconomic shocks. Such assumptions require little defense. They merely require that scientific or other exogenous error in measuring the signal is idiosyncratic, and unrelated to business cycle fluctuations. In other words, taken together (3.2) and (3.3) implicitly define what types of data are good signals: They must be at once related to economic activity in a known and useful way, and measurement error must not suffer from endogeneity. The analogy to valid instruments is immediate. If either of these assumptions is violated, changes in y^*_{nt} may not be reflected in the signal, which would in this case become unreliable.

The motivation of combined measure construction is that traditionally reported observable output data y_{nt} is erroneous.

$$y_{nt} = y^*_{nt} + u^*_{nt} \tag{3.4}$$

Unlike signal measurement error ε^s_{nt}, however, output measurement error u^*_{nt} is allowed to be serially correlated.

$$u^*_{nt} = \rho_u u^*_{nt-1} + \varepsilon^u_{nt} \tag{3.5}$$

$0 < |\rho_u| < 1$, $\varepsilon^u_{nt} \sim IWN(0, \sigma^2_u)$ for $0 < \sigma_u < \infty$, $\{\varepsilon^u_{nt}\}$ has finite fourth moments, and $\{\varepsilon^u_{nt}\}$ is independent of $\{\varepsilon^s_{nt}\}$ and $\{\varepsilon^y_{nt}\}$. In words, ε^u_{nt} may be correlated across areas, though not time. It is not correlated with either macroeconomic shocks, or signal measurement error.

Allowing for u^*_{nt} to be serially correlated is the second significant way in which this paper departs from the assumptions of Henderson et al. The purpose of this generalization is both statistical and structural. From a strictly statistical perspective, the basic model (3.1) and (3.2), while theoretically elegant, can for the same reason not be expected to account for all comovements in the data. Serial correlation in measurement error enables the model to more closely match observed experience, without confounding parsimony in specification. This approach has a long tradition in economic modeling (Sargent, 1989). Finally, from this perspective $\{\varepsilon^u_{nt}\}$ is defined outside the scope of the present model, such as from an omitted structural shock. This supports its independence from $\{\varepsilon^s_{nt}\}$ and $\{\varepsilon^y_{nt}\}$.

From a more substantive structural perspective, since u^*_{nt} is persistent, its process also allows for systematic human-related intervention in data reporting. Consider the situation in which an authority for area n designates an output target.[12] This target is reasonably derived from macroeconomic fundamentals about what rule of motion output has: $\tau_{nt} = \rho_y \tau_{nt-1} + \varepsilon^y_{nt} + \varepsilon^\tau_{nt}$ and $\varepsilon^\tau_{nt} \sim IWN(0, \sigma^2_\tau)$ for $0 < \sigma_\tau < \infty$. In words, the target, in units percentage change less-means, follows the same rule of motion as output, with some error. This would reflect the will of an authority with knowledge about the true structure of the economy, but with expectations about the effectiveness of policies outside of the scope of (3.1). In opposition or party to this

[12] With respect to China, this "authority" is equivalent to the nationally governing Communist Party.

authority, assume that there is an output reporter who manipulates output data y_{nt} exactly to the extent such that it achieves this target in each period: $u^*_{nt} = \tau_{nt} - y^*_{nt}$.[13] Substituting yields $u^*_{nt} = \rho_y u^*_{nt-1} + \varepsilon^{\tau}_{nt}$. This is identically the assumed rule of motion for reporting error (3.5) under the cross-equation restrictions $\rho_u = \rho_y$ and $\varepsilon^u_{nt} = \varepsilon^{\tau}_{nt}$. From this perspective, $\{\varepsilon^u_{nt}\}$ is purely institutional in nature, supporting its independence from $\{\varepsilon^s_{nt}\}$ and $\{\varepsilon^y_{nt}\}$.

In sum, if one had reason to believe: That there is an output target which is in part based upon the economy's true output evolution; that true output were persistent; and that reporters of output were incentivized to meet this target, then these points would logically *require* that u^*_{nt} were persistent. Thus (3.1) and (3.5) jointly allow the model to be general enough to capture a range of pertinent political behavior not allowable under classical measurement error. An example is the endogenized targeting-misreporting framework just described. Yet, the model at the same time takes no stance as to whether the source of persistence in u^*_{nt} is so-called statistical or structural, as defined above. Thus, it is appropriate in either case.

3.2 Combined measure

(3.1)-(3.5) encapsulate the model. In total, there are six structural parameters, collected in

$$\underset{6\times 1}{\theta} = (\underbrace{\beta, \sigma}_{\text{Signal}}, \underbrace{\rho_y, \sigma_y}_{\text{Output}}, \underbrace{\rho_u, \sigma_u}_{\text{Error}})' \tag{3.6}$$

The structural parameters may be partitioned into two which are signal-specific, and four which are not. Of those which are not, two depend upon the y-evolution of latent output, and two upon the u-evolution of output measurement error. The identifiability of θ and a consistent estimator $\widehat{\theta}$ will be discussed in the following section. Given any consistent estimator, and known relationship between signal and latent output (3.2) and (3.3), one may construct a proxy for growth. It is the linear projection

$$\widehat{z}_{nt} = \left(1/\widehat{\beta}\right) s_{nt} \tag{3.7}$$

The proxy $\widehat{z}_{nt}$ may be used to construct a composite estimate of output. Such a combined measure is the weighted average $x_{nt} = (1-\phi)y_{nt} + \phi\widehat{z}_{nt}$. Creating an optimal combined measure $\widehat{x}_{nt}$ means

[13]With respect to China, this "reporter" is equivalent to a provincial or other subnational level figurehead who vies for advancement.

choosing the proxy loading ϕ accordingly. The appropriate choice is a consistent estimator $\widehat{\phi}$ for ϕ which minimizes the mean squared error of x_{nt}. The mean squared error is

$$V(\phi) = E\left[(1-\phi)y_{nt} + \phi\widehat{z}_{nt} - y^*_{nt}\right]^2 = E\left[(1-\phi)u^*_{nt} + (\phi/\beta)\varepsilon^s_{nt}\right]^2 = (1-\phi)^2\frac{\sigma^2_u}{1-\rho^2_u} + \frac{\phi^2}{\beta^2}\sigma^2$$

The first order condition of $V(\phi)$ evaluated at the estimator is

$$V'(\phi)\big|_{\theta=\widehat{\theta}} = 0 = -2(1-\widehat{\phi})\frac{\widehat{\sigma}^2_u}{1-\widehat{\rho}^2_u} + \frac{2\widehat{\phi}}{\widehat{\beta}^2}\widehat{\sigma}^2$$

Simply rearranging yields the optimal loading.

$$\widehat{\phi} = \frac{\widehat{\sigma}^2_u}{1-\widehat{\rho}^2_u} \Bigg/ \left(\frac{\widehat{\sigma}^2_u}{1-\widehat{\rho}^2_u} + \frac{\widehat{\sigma}^2}{\widehat{\beta}^2}\right) \tag{3.8}$$

From this, we may finally define the optimal combined measure of output.

$$\widehat{x}_{nt} = (1-\widehat{\phi})y_{nt} + \widehat{\phi}\widehat{z}_{nt} \tag{3.9}$$

$\widehat{\phi} \in [0,1]$ is required of its definition as a weighting. A necessary and sufficient condition is $|\widehat{\rho}_u| \in [0,1]$. In the extreme case that $\widehat{\rho}_u = 0$ and output measurement error is not serially correlated, both traditional output and the signal are potentially useful. When this is so, the magnitude of the loading depends upon (1) the elasticity of the signal with respect to output, and (2) the relative magnitudes of measurement errors. Conversely, in the extreme case that $\widehat{\rho}_u = 1$ and output measurement error is a random walk, then reported data is not reliable, and the combined measure is identically the proxy. The previous assumptions restrict $|\widehat{\rho}_u| \xrightarrow{p} |\rho_{u0}| \in (0,1)$ and $\widehat{\beta} \xrightarrow{p} \beta_0 \neq 0$ for consistent estimators. $\xrightarrow{p}$ denotes convergence in probability and subscript-0 denotes population value. In any such admissible case, $\widehat{\phi} \xrightarrow{p} \phi_0 \in (0,1)$ and the economic implication is some nontrivial convex combination of the two extreme outcomes.

3.3 Multiple signals

The analysis thus far has concerned itself with a lone signal s_{nt}. As many distinct combined measures may be computed as there are signals available. However, should there in fact be many signals available, one may naturally wish to instead compute a combined measure from all signals, jointly. Let us assume that there are $i = 1, \ldots, S$ signals under consideration. In terms of these, define the following vectors of measured signals, true values, and errors.

$$\underset{S\times 1}{s_{nt}} = \begin{bmatrix} s_{nt}(1) & \ldots & s_{nt}(S) \end{bmatrix}' \qquad \underset{S\times 1}{s^*_{nt}} = \begin{bmatrix} s^*_{nt}(1) & \ldots & s^*_{nt}(S) \end{bmatrix}' \qquad \underset{S\times 1}{\varepsilon^s_{nt}} = \begin{bmatrix} \varepsilon^s_{nt}(1) & \ldots & \varepsilon^s_{nt}(S) \end{bmatrix}'$$

Henceforth, parenthetic (i) denotes parameters and variables which are signal i-specific. The system equivalent of the signal measurement equation (3.3) is again written,

$$s_{nt} = s^*_{nt} + \varepsilon^s_{nt} \tag{3.10}$$

but now, s_{nt}, s^*_{nt}, and ε^s_{nt} are vectors. In the case of a lone signal $S = 1$, there are 6 structural parameters in θ (3.6). The first two (β, σ) are signal-dependent, while the latter four $(\rho_y, \sigma_y, \rho_u, \sigma_u)$ are signal-independent. The system equivalent has $S + S(S+1)/2$ structural parameters which are signal-dependent for each $i = 1, \ldots, S$, while the latter four are mutual amongst signals. Thus, there are $n_\theta = S + S(S+1)/2 + 4$ total structural parameters for $S \geq 1$ signals.

$$\underset{n_\theta \times 1}{\theta} = (\underbrace{\beta', \sigma'}_{S\text{ Signals}}, \underbrace{\rho_y, \sigma_y}_{\text{Output}}, \underbrace{\rho_u, \sigma_u}_{\text{Error}})' \tag{3.11}$$

$$\underset{S\times 1}{\beta} = \begin{bmatrix} \beta(1) & \ldots & \beta(S) \end{bmatrix}' \qquad \underset{S(S+1)\times 1}{\sigma} = \text{vech}(L) \qquad \underset{S\times S}{L} = \text{chol}(\Sigma) \qquad \underset{S\times S}{\Sigma} \equiv E(\varepsilon^s_{nt}\varepsilon^{s'}_{nt})$$

vech$(\cdot)$ is the operator which selects the $S(S+1)/2$ unique elements on and below the principal diagonal. chol$(\cdot)$ is the lower-left Cholesky decomposition so that $LL' = \Sigma$. Σ is allowed to have possibly nonzero off-diagonal elements for the fact that signals may be "reconciled" to one another. Explicitly, if certain candidate signals are manipulated in tandem to veil output data manipulation, then their measurement errors should be expected to be mutually correlated. But at the same time, that these correlations are zero is simply a special case, and such speculations need not be entertained. Finally, note that in the special case that $S = 1$, $n_\theta = 6$, and we have the previous lone signal model setup.

Given a consistent estimator $\widehat{\theta}$, we also have S independent proxies for economic growth.

$$\underset{S\times 1}{\widehat{z}_{nt}} = \text{diag}(\widehat{\beta})^{-1} s_{nt} \tag{3.12}$$

Recall, the individual elasticities are nonzero by assumption of the signals' relevance. Therefore, $\text{diag}(\widehat{\beta})$, the $S\times S$ square diagonal matrix with the elements of β in order on its principal diagonal, is always invertible.

Finally, we wish to utilize these S proxies $\widehat{z}_{nt}$ to construct a joint composite estimate of growth. Such a joint combined measure is the weighted average $x_{nt} = (1-\phi' 1_S)y_{nt} + \phi'\widehat{z}_{nt}$ where ϕ is an $S\times 1$ vector of loadings and 1_S is an $S\times 1$ vector of 1's. Creating an optimal combined measure $\widehat{x}_{nt}$ means choosing the proxy loadings ϕ accordingly. The appropriate choice is a consistent estimator $\widehat{\phi}$ for ϕ which minimizes the mean squared error of x_{nt}. The mean squared error is

$$V(\phi) = E\left[(1-\phi' 1_S)y_{nt} + \phi'\widehat{z}_{nt} - y^*_{nt}\right]^2 = (1-\phi' 1_S)^2 \frac{\sigma_u^2}{1-\rho_u^2} + \phi'\text{diag}(\widehat{\beta})^{-1}\Sigma\text{diag}(\widehat{\beta})^{-1}\phi$$

The first order condition of $V(\phi)$ evaluated at the estimator is

$$\underset{1\times S}{\left.\frac{\partial V(\phi)}{\partial \phi'}\right|_{\theta=\widehat{\theta}}} = 0_{1\times S} = -2(1-\widehat{\phi}' 1_S)1_S' \frac{\widehat{\sigma}_u^2}{1-\widehat{\rho}_u^2} + 2\widehat{\phi}'\text{diag}(\widehat{\beta})^{-1}\widehat{\Sigma}\text{diag}(\widehat{\beta})^{-1}$$

Simply rearranging yields the optimal loadings.

$$\underset{S\times 1}{\widehat{\phi}} = \left[1_S 1_S' \frac{\widehat{\sigma}_u^2}{1-\widehat{\rho}_u^2} + \text{diag}(\widehat{\beta})^{-1}\widehat{\Sigma}\text{diag}(\widehat{\beta})^{-1}\right]^{-1} \times 1_S \frac{\widehat{\sigma}_u^2}{1-\widehat{\rho}_u^2} \tag{3.13}$$

$\widehat{\phi}' 1_S \in [0,1]$ is required of ϕ's definition as a weighting. A necessary and sufficient condition, as in the lone signal case, is $|\widehat{\rho}_u| \in [0,1]$. Intuitively, when $S=1$, $\widehat{\phi}$ reduces to the lone signal weighting (3.8). However, (3.13) is not equivalent to a vector of (3.8) computed signal-by-signal for $S>1$. In that case $\widehat{\phi}' 1_S \in [0,1]$ is not generally satisfied. The optimal multiple signal combined measure is

$$\widehat{x}_{nt} = (1-\widehat{\phi}' 1_S)y_{nt} + \widehat{\phi}'\widehat{z}_{nt} \tag{3.14}$$

4 Methodology

Aside from a generalization to multiple signals, the construction of combined measures described in the previous section is in many ways similar to the treatment of Henderson et al. (2012). The lynchpin in both cases is the estimation of the loadings ϕ. However, there are subtle practical points yet to be considered. For example, a primitive assumption underlying the existence of a consistent estimator $\widehat{\phi}$ is that all elements of θ are identifiable. In Henderson et al., the analogous parameter to σ_u – the variability of output reporting error – is not identifiable. The authors fix this parameter to a constant, parameterizing the value by-area using A-through-D country data quality grades provided in the Penn World Tables.

But fixing this parameter produces bias. At hand is the fact that conditioning is not a good practice when the conditioning information – like the coarse Penn World Table rankings in particular – is imprecise. Let us consider a concrete example. Say that Country A is characterized by population values $\beta = 1$, $\sigma_0 = 10/9$, $\rho_{u0} = 1/10$, and $\sigma_{u0} = 1$ for some lone signal. Then $\phi_0 = 1/2$. But also say that Country A has received the same data quality grade as Country B, which has $\sigma_{u0} = 1/2$ in population. Inherently, categorical rankings provide only such blunted characterizations of true data quality, regardless of the realized magnitude of differences within-bins. If σ_u is fixed to $\overline{\sigma}_u = 1/2$ for both Country A and B, estimating ϕ conditionally for Country A gives $\widehat{\phi}|\overline{\sigma}_u \xrightarrow{p} 1/3$. The proxy is asymptotically underweighted, despite being entirely consistent with categorical rankings. Unless all countries within each grade have exactly the same population values, an assumption which is plainly unreasonable, then such unquantifiable biases are a forgone conclusion.

Moreover, existing panel estimates of combined measures have primarily exploited a large N cross sectional dimension. This focus is the result of the small time samples T previously under consideration. For example, the DMSP-OLS luminosity panel is currently only annual from 1992 to 2013, making time series analysis for single areas difficult. Nonetheless, as solidified by the analysis in Section 2, inevitable locational heterogeneity in production technology and input mix imply that the elasticity β should not be expected to be the same from country to country, or even within. In this case, overzealously pooled estimators are not consistent for large N, regardless. Given that the time series dimension available for NO_X and other relevant signals are considerably more generous, the applicability of fixed N large T asymptotics is in that context more promising. But this requires the development of a new estimator, and associated proofs of large T consistency and asymptotic normality.

This section first establishes that the model observables have "structural" dynamic panel data representation. Amongst other ramifications, this structure implies that the errors are serially correlated. Consequently, an instrumental variables generalization of the within estimator typically utilized to estimate dynamic panel models becomes useful. Inferring latent growth, however, also requires a second step of recovering the structural parameters from these estimates. The nuisance parameters which arise in this nonlinear mapping elucidate why any attempt to decipher latent growth from purely linear inferential techniques is likely to yield misleading conclusions. Finally, this ultimately computationally efficient estimator also expedites the bootstrapping of small T sample bias correction, confidence intervals, Wald statistics, and error bands for the combined measure. These statistics help alleviate biases intrinsic to dynamic panel data estimation when N approaches ∞ faster than T, and sample biases incurred for fixed N small T. Inasmuch, they are pivotal to subsequent empirical analysis.

4.1 Identification and estimation of reduced form

The model (3.1)-(3.5) contains both observable and unobservable variables, implying state space representation when utilizing any lone signal. Proposition 1, proven in Appendix C, establishes that this state space also yields parsimonious representation for the observables given any $S \geq 1$ set of signals.

Proposition 1. *(Reduced Form Representation.) The observables have the representation*

$$Y_{nt} = X_{nt}\Psi + V_{nt} \tag{4.1}$$

for $X_{nt} = \left(Y'_{nt-1} \otimes I_{S+1}\right) R$ *is* $(S+1) \times (S+2)$, $Y_{nt} = \begin{bmatrix} y_{nt} & s'_{nt} \end{bmatrix}'$ *is* $(S+1) \times 1$, R *an* $(S+1)^2 \times (S+2)$*-dimensional zero-one selection matrix,* Ψ *a* $(S+2) \times 1$ *vector function of both the structural parameters* θ *and* S *nuisance parameters contained in* λ *which sum to 1*

$$\Psi(\theta; \lambda) = \begin{bmatrix} \rho_y & \rho_u & \psi' \end{bmatrix}' \tag{4.2}$$

$$\underset{S\times 1}{\psi(\theta;\lambda)} = (\rho_y - \rho_u)\begin{bmatrix}\frac{\lambda_1}{\beta(1)} & \dots & \frac{\lambda_S}{\beta(S)}\end{bmatrix}' \tag{4.3}$$

$$\underset{S\times 1}{\lambda} = \begin{bmatrix}\lambda_1 & \dots & \lambda_S\end{bmatrix}'; \qquad \lambda' 1_S = 1 \tag{4.4}$$

and V_{nt} *a* $(S+1)\times 1$ *vector MA(1) error with variance-covariance matrix*

$$\Omega(\theta;\lambda) = E(V_{nt}V_{nt}') \tag{4.5}$$

for Ω *a nonlinear but closed form function of* θ *and* λ.

Equation (4 .1) is interpretable from the time series dimension as by-region n restricted VARMA(1,1) representation. Exclusion restrictions are embodied in R and cross-equation restrictions in Ψ and Ω. This representation is useful in that it is stated entirely independently of the unobserved states. Yet, serial correlation in the errors $\{V_{nt}\}$ is directly attributable to the model's inherent dependence on the unknown status of output and reporting error. The existence of such latent states precludes finite order VAR representation, for example, in all but special circumstances (Ravenna, 2007).

From a panel perspective, however, this representation is more naturally interpreted as a generalization of dynamic panel data (DPD). Specifically, this is vector DPD with serially correlated errors and structurally founded parametric restrictions. Aside from vector notation, structural DPD has several more substantive distinguishing characteristics from the more familiar DPD. First, the object of interest from the perspective of identification, estimation, and inference is the structural parameters, not the reduced form. However, it is only in the case that $S = 1$ when $\lambda_1 = 1$ is known a-priori. So, any attempt to identify the structural parameters from multiple signals without accounting for λ will always prove fruitless. Second, the fact that $\{V_{nt}\}$ is serially correlated means that commonly utilized DPD estimators which assume otherwise will not yield consistent estimates for Ψ itself, regardless. Third, the application of interest suggests large-T asymptotic results are at least as important as large-N results, since locational heterogeneity in production technology will generally imply distinct elasticities of a given signal with respect to income. This hinders the applicability of GMM estimators which presume N grows faster than T (Holtz-Eakin et al., 1988; Arellano and Bond, 1991), or bias-correction methodologies which rely on large N and T (Hahn and Kuersteiner, 2002; Bai, 2009).

So, while we are ultimately interested in estimation and inference for θ, such an analysis remains yet impossible unless it is first possible to identify Ψ and Ω. Furthermore, estimators which seek to recover θ directly, such as maximum likelihood, may be numerically cumbersome, a feature which will subsequently prove undesirable with respect to bootstrap statistics. These facts jointly suggest that the most pragmatic approach to addressing the stated concerns with respect to structural DPD is to first consider the reduced form parameters Ψ and Ω as estimable parameters unto themselves. Then, the more intricate problem of recovering θ may be considered. Proposition 2, proven in Appendix D, extends the familiar within (covariance) estimators from the DPD setting to the structural DPD case.

Proposition 2. *(Estimation of Reduced Form.) The instrumental variables estimator*

$$\underset{(S+2)\times 1}{\widehat{\Psi}} = \left[\sum_{n=1}^{N} \widehat{Z}_n' \widehat{X}_n\right]^{-1} \sum_{n=1}^{N} \widehat{Z}_n' \widehat{Y}_n \tag{4.6}$$

$$\underset{T(S+1)\times 1}{\widehat{Y}_n} = \begin{bmatrix}\widehat{Y}_{n3}' & \dots & \widehat{Y}_{nT+2}'\end{bmatrix}' \qquad \underset{T(S+1)\times(S+2)}{\widehat{X}_n} = \begin{bmatrix}\widehat{X}_{n3}' & \dots & \widehat{X}_{nT+2}'\end{bmatrix}'$$

$$\underset{T(S+1)\times(S+2)}{\widehat{Z}_n} = \begin{bmatrix}\widehat{X}_{n2}' & \dots & \widehat{X}_{nT+1}'\end{bmatrix}' \qquad \widehat{X}_{nt} = (\widehat{Y}_{nt-1}' \otimes I_{S+1})R$$

where $\{\widehat{Y}_{nt}\}$ is the sample analogue to $\{Y_{nt}\}$, is consistent for large T regardless of N

$$\widehat{\Psi} \xrightarrow{p} \Psi_0 \text{ as } T \to \infty \tag{4.7}$$

and asymptotically normal with bias depending on the relative rates of increase of T and N

$$\sqrt{TN}(\widehat{\Psi} - \Psi_0) \xrightarrow{d} N(b(T), Avar(\widehat{\Psi})) \tag{4.8}$$

$b(T) \to 0$ necessarily if N fixed $T \to \infty$, but possibly not if both N and T are increasing. Consistent estimators for $Avar(\widehat{\Psi})$ and Ω are

$$\widehat{Avar}(\widehat{\Psi}) = \left[(TN)^{-1}\sum_{n=1}^{N} \widehat{Z}_n' \widehat{X}_n\right]^{-1} R' \left(\widehat{\Sigma}_Y \otimes \widehat{\Omega}\right) R \left[(TN)^{-1}\sum_{n=1}^{N} \widehat{X}_n' \widehat{Z}_n\right]^{-1} \tag{4.9}$$

$$\widehat{\Sigma}_Y = (TN)^{-1} \sum_{n=1}^{N} \sum_{t=1}^{T} \widehat{Y}_{nt} \widehat{Y}_{nt}' \qquad \widehat{\Omega} = (TN)^{-1} \sum_{n=1}^{N} \sum_{t=3}^{T+2} \widehat{V}_{nt} \widehat{V}_{nt}'$$

for $\widehat{V}_{nt} = \widehat{Y}_{nt} - \widehat{X}_{nt}\widehat{\Psi}$ *the residuals from the estimator* $\widehat{\Psi}$.

Proposition 2 represents an instrumental variables extension of Theorem 1 of Alvarez and Arellano (2003), which investigates the within estimation of scalar DPD models with independent errors. The estimation of Ψ here is possible despite serial correlation in errors simply by using lagged observables as instruments, similarly to Anderson and Hsiao (1981). The first lag is considered, as it is the valid instrument which results in the largest time sample. The incidental parameter bias $b(T)$ which arises in the asymptotic distribution is an artifact of the within transformation necessary to obtain $\{\widehat{Y}_{nt}\}$, as the data contain fixed effects (Neyman and Scott, 1948; Nickell, 1981). This bias may not converge to zero if N is not fixed. Yet, it will necessarily converge to zero for fixed N large T.

4.2 Identification and estimation of structural parameters

When attempting to deduce true growth from signals, the natural inclination of an analyst may be to include as many signals as possible in a simple reduced form linear analysis. For example, in the form of a vector autoregression. However, Proposition 1 implies that *no* linear inferential approach based on multiple signals necessarily identifies the elasticities β necessary to make inferences about the unobserved state y_{nt}^*. This is due to the existence of nuisance parameters $\{\lambda_i\}$ when $S > 1$. Specifically, one must identify nuisance parameters separately of elasticities. The only exception to this rule is when there is but $S = 1$ signal, as in Henderson et al.'s luminosity study.

With this in mind, one may now consider the necessarily nonlinear problem of identification and estimation of θ when $S \geq 1$ given the estimators and restrictions $\widehat{\Psi}$, $\widehat{\Omega}$, and $\lambda' 1_S = 1$. The easiest way to begin any nonlinear identification exercise is to verify that necessary order conditions are satisfied. Specifically, in order to separately identify θ and λ, there must be at least as much information in the $n_\theta + S$ vector of reduced form restrictions, π.

$$\underset{(n_\theta+S)\times 1}{\pi} = \begin{bmatrix} \lambda' 1_S & \Psi' & \mathrm{vech}(\Omega)' \end{bmatrix}' \tag{4.10}$$

Table 3: Exact identification of structural θ and nuisance λ parameters.

Restriction	No. Elements	λ S	β S	ρ_y 1	ρ_u 1	σ_y 1	σ_u 1	σ $S(S+1)/2$	Total $n_\theta + S$
$\lambda' 1_S$	1	✓							
Ψ	$S+2$	✓	✓	✓	✓				
$\text{vech}(\Omega)$	$(S+1)(S+2)/2$	✓	✓	✓	✓	✓	✓	✓	
Total	$n_\theta + S$								

One way to cogently assess the dimension of either set of parameters is to enumerate them in a tableau. Such an analysis is presented in Table 3. As described by this table, there are exactly as many structural and nuisance parameters as there are restrictions to be exploited for their identification. Therefore, the order condition is just satisfied for any $S \geq 1$. Appendix E.1 establishes that when $S = 1$, the mapping π is in fact simple enough to invert analytically, which inasmuch guarantees θ is globally identified. This feature also means that an efficient indirect least squares estimator for θ may be written in terms of the inverse mapping.

$$S = 1: \ \widehat{\theta} = \pi^{-1}\left(\begin{bmatrix}\widehat{\Psi}' & \text{vech}(\widehat{\Omega})'\end{bmatrix}'\right) \tag{4.11}$$

When $S > 1$, π becomes too complicated to invert analytically. But Appendix E.1 also establishes that one may still compute analytically the square Jacobian

$$\underset{(n_\theta+S)\times(n_\theta+S)}{J(\theta;\lambda)} = \frac{\partial \pi'}{\partial \begin{bmatrix}\theta' & \lambda'\end{bmatrix}} \tag{4.12}$$

As will be substantiated numerically in the following subsection, this matrix is typically full column rank. This ensures point local identifiability of the structural parameters (Rothenberg (1971)). Furthermore, despite the fact that π^{-1} is infeasible for $S > 1$, it is yet possible to compute an asymptotically efficient 2-step estimator on the basis of the Jacobian's inverse.

$$S > 1: \ \begin{bmatrix}\widehat{\theta} \\ \widehat{\lambda}\end{bmatrix} = \begin{bmatrix}\widehat{\theta}^* \\ \widehat{\lambda}^*\end{bmatrix} + J(\widehat{\theta}^*, \widehat{\lambda}^*)^{-1}\left(\begin{bmatrix}1 \\ \widehat{\Psi} \\ \text{vech}(\widehat{\Omega})\end{bmatrix} - \begin{bmatrix}\widehat{\lambda}^{*\prime} 1_S \\ \Psi(\widehat{\theta}^*, \widehat{\lambda}^*) \\ \text{vech}(\Omega(\widehat{\theta}^*, \widehat{\lambda}^*))\end{bmatrix}\right) \tag{4.13}$$

where $\widehat{\theta}^*$ and $\widehat{\lambda}^*$ are the signal-by-signal estimators for θ and λ constructed from each individual signal's consistent estimator (4.11). Appendix E.2 pedantically details the steps to be followed in constructing this two-step estimator and establishes its asymptotic equivalence to the infeasible efficient estimator based on inverting the mapping π directly.

4.3 Inference

When $S = 1$, Ψ is 3-dimensional and depends upon three structural parameters – β, ρ_y, and ρ_u – and no nuisance parameters. As a result, confidence intervals for the indirect least squares estimator (4.11) for these parameters can easily be computed using $\widehat{\text{Avar}}(\widehat{\Psi})$ by the delta method. Yet, in the more general case one wishes to compute confidence intervals for the entire vector of structural parameters θ, or if $S > 1$, such familiar results are not applicable. This is due to the fact that such statistics would also depend upon the non-Gaussian asymptotic distribution of $\widehat{\Omega}$. Moreover, analytical corrections for the incidental parameter bias $b(T)$ do not exist.

The fact that $\widehat{\theta}$ is a computationally efficient estimator for all $S \geq 1$ does however make bootstrapping such statistics feasible. Recall, there is inherently bias in the DPD estimates if N is increasing with T. While there is no bias for fixed N large T, there will also likely be bias for fixed N small T. In order to come to grips with the magnitude of these biases, consider the following values for a hypothetical model with $S = 1$:

$$\beta = 3, \quad \sigma = 1\text{e} - 3, \quad \rho_y = 0.7, \quad \sigma_y = 1\text{e} - 3, \quad \rho_u = 0.5, \quad \sigma_u = 1\text{e} - 3. \tag{4.14}$$

Using (4.14) as data-generating values, the bias in the indirect least squares estimator for θ with variable N and T is computed by Monte Carlo experiment. The results are summarized by Table 4. Recall, the annual sample utilized in this paper has $N = 31$ and $T = 16$ while the quarterly sample has $T = 31$. Samples with cross-sectional dimension of $N = 10$ and lower are investigated in these Monte Carlo experiments, since as discussed, it is desirable to partition provinces, municipalities, and autonomous regions into as many distinct tranches as possible. In the entire range of potential panel dimensions to be utilized, $\widehat{\beta}$ is consistently downward-biased by approximately 10% of its true value. So, the suggestion adhered to henceforth is to utilize the

Table 4: Small sample bias.

	$N=1$		$N=3$		$N=6$		$N=10$	
	$T=15$	$T=30$	$T=15$	$T=30$	$T=15$	$T=30$	$T=15$	$T=30$
β	-0.26	-0.35	-0.37	-0.41	-0.39	-0.36	-0.39	-0.33
σ	0	0	0	0	0	0	0	0
ρ_y	-0.08	-0.03	0	0	0	0	0	0
σ_y	0	0	0	0	0	0	0	0
ρ_u	-0.25	-0.08	-0.08	-0.07	-0.07	-0.06	-0.06	-0.04
σ_u	0	0	0	0	0	0	0	0
ϕ	0.44	-0.07	-0.44	-0.16	-0.16	-0.16	-0.17	-0.15

bias-corrected estimator,

$$\widehat{\theta}^{\star} = 2\widehat{\theta} - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{(b)} \tag{4.15}$$

with bootstrap draws $\{\widehat{\theta}^{(b)}\}$ (see Appendix F).

Next, we wish to numerically examine the coverage probabilities of either asymptotic or bootstrap confidence intervals for the bias corrected estimator $\widehat{\theta}^{\star}$ when $S = 1$. Using again (4.14) as data-generating values and $S = 1$, asymptotic 95% confidence intervals are computed for β, ρ_y, and ρ_u over a range of N and T in the top pane of Table 5. In all cases, β is significantly under-covered. It is impossible to compute asymptotic confidence intervals for the remaining structural parameters, or the weighting ϕ, with standard inferential tools.

In order to investigate the possibility that bootstrap confidence intervals perform better, Monte Carlo experiments to compute the bootstrap confidence interval's coverage probability are carried out using the methodology described by Horowitz (2001).[14] Results are described in the bottom pane of Table 5. The results indicate that for time series samples of $T = 15$, roughly the magnitude considered in the annual sample in this paper, a cross sectional sample of minimally $N = 10$ is required to obtain correct coverage probabilities for all parameters of interest. However, if the time series dimension is expanded to $T = 30$, roughly the magnitude considered in the quarterly sample in this paper, a cross sectional sample of just $N = 6$ is required to obtain correct coverage. In either case, these results indicate that bootstrap confidence intervals

[14]See Appendix F for the computation of confidence intervals. Following Horowitz's design, for each of 1000+ Monte Carlo draws, a bootstrap confidence interval, requiring another 1000+ draws, is computed. The dimensionality of this computation makes it intensive. Code was parallelized across 40 CPUs and took several hours to compile.

Table 5: Coverage probability of 95% confidence interval.

Asymptotic.

	$N=1$		$N=3$		$N=6$		$N=10$	
	$T=15$	$T=30$	$T=15$	$T=30$	$T=15$	$T=30$	$T=15$	$T=30$
β	11	6	4	2	1	1	1	1
ρ_y	92	93	91	92	85	88	80	79
ρ_u	93	94	94	93	87	91	84	89

Bootstrap.

	$N=1$		$N=3$		$N=6$		$N=10$	
	$T=15$	$T=30$	$T=15$	$T=30$	$T=15$	$T=30^*$	$T=15^*$	$T=30^*$
β	95	97	99	98	98	97	96	95
σ	96	98	98	98	98	98	98	98
ρ_y	92	94	96	98	99	99	99	99
σ_y	96	98	99	98	98	96	97	96
ρ_u	67	71	77	87	87	94	93	97
σ_u	91	92	92	95	96	96	96	98
ϕ	92	93	93	96	95	95	95	96

Notes: * indicates correct coverage probability.

are useful in this application; they not only have good coverage probabilities, but may be used to compute uncertainty for *all* parameters of interest. In the ensuing empirical analysis, these results are used to inform bin sizes for estimation, in which we wish to keep N as small as possible for each T.

Finally, Appendix F also details the computation of percentile confidence interval for the computed combined measure series $\{\widehat{x}_{nt}\}$ itself. These are henceforth known as *error bands.* The key economic question of this paper is whether officially reported output statistics may be validated using other signals of growth. Naturally, in any given sample, the combined measure $\widehat{x}_{nt}$ and reported figure y_{nt} may differ. Is this evidence that officially reported statistics may not be validated? No. Not just the combined measure, but its entire sampling distribution, is the object of interest: Is a given purported difference between the combined measure and reported output, $\widehat{x}_{nt} - y_{nt}$, *statistically significant*? Level α error bands provide a range in which actual output figures are likely to lie, on the basis of signals of economic growth. If reported GDP numbers fall outside of these bands, they may be regarded as statistically improbable. Precisely, there is a $1-\alpha$

probability that this conclusion is incorrect, and that reported figures are indeed representative of true growth (Type I error). Error bands are thereby fundamental to understanding when reported output figures are feasible. They are the main statistical object of interest for drawing substantive economic conclusions in the following section.

5 Results

The analysis centers on two trying periods in the recent economic history of China: The Asian Financial Crisis period of 1997-1999, and the Global Financial Crisis and Great Recession period of 2007-2009. The central question is whether signals of economic growth – such as luminosity, NO_X emissions, or indices of industrial production – may be used to validate officially reported figures at the subnational level during either of these periods. In order to critically consider this question, this section applies the model described in Section 3 and methodology developed in Section 4 to the datasets described in Section 2.

5.1 The Asian Financial Crisis

Recalling the discussion of Section 2, the annualized data set, spanning from 1993-2008 ($T = 16$) and across all $N = 31$ regions, is utilized. The previous analysis has indicated that pooling the estimates across all regions is likely to lead to misleading conclusions; ideally, estimates would be computed on a region-by-region basis. But at the same time, the Monte Carlo experiment summarized by Table 5 indicates that no cross sectional dimension less than $N = 10$ is likely to lead to correct coverage probabilities with this time dimension.

For this reason, the $N = 31$ subnational regions are to be separated into two groups of 10 and one of 11 before estimation. For each of the three group-dependent estimators to be consistent, the structural parameters θ must be in common of all regions within each bin. As an objective means of choosing these groupings, one may utilize the preliminary by-region annual NO_X emissions estimated elasticities $\widehat{\beta}$ listed in Figure 6 to provide an ordering. The groupings implied by these preliminary estimates are depicted in Figure 8 (a); Group 1 ($N = 10$) contains highest elasticities, Group 3 ($N = 10$) contains smallest elasticities, and Group 2 ($N = 11$) contains mid-range. Note, there is no obvious geographic or economic dependence amongst members in each group, outside of the elasticity estimate. Similar observations hold were we to use the preliminary by-region annual luminosity estimated elasticities $\widehat{\beta}$ also listed in Figure 6.

Using these groupings, bias-corrected estimates for the structural parameters θ and weightings ϕ are computed. Recall, ϕ is in some sense the primary parameter of interest, since larger ϕ estimates indicate the given proxy is useful in combined measure construction. Signal-by-signal estimates for ϕ with 95% bootstrap confidence intervals are listed in Table 6 ($S = 1$ columns). Estimates for each structural parameter in θ along with confidence intervals are listed in Appendix Tables A.3-A.4. Relatively large confidence intervals in all cases disallow us from attributing statistical significance to ϕ or the elasticity β for either luminosity or NO_X emissions. But can we take a stance on which signal marginally preferred by the data? Table 6 also presents estimates using all signals, jointly, with confidence intervals not reported ($S = 5$ columns).[15] There are two important observations. First, the data tend to prefer luminosity in all groups, with a weighting as high as 1 in Group 3. This implies that the luminosity data is certainly useful on an annualized basis. Second, summing up weighing adds to almost identically 1 in all cases. In other words, even reported output always receives a weighting of zero. It seems that utilizing all signals jointly may tend drown out the information contained in reported figures.

The ultimate purpose of this analysis is to determine which officially reported output data are supported by other signals of growth. Bootstrap error bands, described in Section 4.3 and computed using any signal or signals, provide a formal means for determining this. Figure 9 depicts the error bands computed by-region using the perviously described estimates using luminosity alone. Note, in several instances reported output (black line) escapes the luminosity-based error band (pink shading). This indicates that one may reject the null hypothesis that reported output data is consistent with remotely measured luminosity readings during the given period, with a 5% chance that this rejection is incorrect.

But can we also reasonably conclude that these discrepancies are due to politically motivated misreporting in all instances? It seems that the only reasonable answer here is no. Consider that the validity of reported output is rejected on the basis of luminosity data *before* the financial crisis period (Beijing, Tianjin, and others). This period, 1993-1997, was in fact a period of fantastic

[15]Confidence intervals for $S > 1$ can be computed using a similar bootstrap methodology as for $S = 1$. However, since the use of the multiple signal estimator is to judge which signals are comparatively favored by the data, these are not computed. In addition, the Jacobian J is full column rank, ensuring local identification of the structural parameters.

Table 6: Estimates: Annual sample, 1993-2008.

	Group 1 ($N = 10$)		Group 2 ($N = 11$)		Group 3 ($N = 10$)		Pooled ($N = 31$)	
	$S = 1$	$S = 5$	$S = 1$	$S = 5$	$S = 1$	$S = 5$	$S = 1$	$S = 5$
ϕ_L	-0.40	0.81	0.01	0.99	0.81	1	0.27	1
	(-0.81,1.23)		(-1.82,1.91)		(-0.14,2.38)		(-0.38,1.92)	
ϕ_N	-0.69	6e-4	-1.48	-1e-3	0.22	0	-0.60	0
	(-1.00,0.81)		(-2.16,-0.06)		(-0.19,1.87)		(-0.94,1.03)	
ϕ_F	0.97*	0.18	1.24*	2e-3	-0.79	0	-0.97	0
	(0.97,1.02)		(0.78,2.80)		(-2.03,2.26)		(-1.59,0.68)	
ϕ_E	0.62*	4e-4	-0.80	0	-1.19	0	-0.76	0
	(0.07,1.98)		(-1.28,0.64)		(-2.29,0.37)		(-1.23,0.76)	
ϕ_C	1.18*	0.01	0.95*	2e-4	1.04*	0	1.05*	0
	(0.82,2.77)		(0.93,1.16)		(1.03,1.07)		(0.87,1.82)	

Notes: * Significance at 95% confidence level (confidence interval). Groupings are described by Figure 8 (a). L: Luminosity. N: NO_X emissions. F: Freight volume. E: Electricity generation. C: Cement production.

growth in the Chinese economy. Rather than misreporting, what we are more likely witnessing is few years of particularly rapid decline leading into the recessionary period. In other words, the variation in the remainder of this annual data sample is not sufficient enough to allow one to avoid Type I errors in the beginning of the sample. As additional evidence of this concern, note that reported output figures appear to be too low *during* the crisis period in several regions (Tianjin, Hebei, and others). It seems highly unlikely that regional officials would report output figures which were too low during a crisis period. In other words, the statistical inference implied by this aberration is not consistent with our economic story.

In order to motivate that discrepancies are politically motivated, we minimally must observe that discrepancies between reported figures and error bands occur at politically opportune times: During or subsequent to the crisis period. Is the problem luminosity as a signal? Also no. Similar error bands for NO_X are given in the Appendix, Figure A.5, along with freight, electricity, and cement, in Figures A.6 - A.8 respectively. In all cases, we arrive to similar unreasonable conclusions as with luminosity itself: That reported output was "too high" during the boom years before the Crisis, and "too low" during. To focus this line of inquiry, Figure 10 presents error bands using each individual signal across the four major municipalities in the sample: Beijing, Tianjin, Shanghai, and Chongqing. In essentially all cases, the estimation suggests the same conclusion: That annual data is not volatile enough to pinpoint evidence of data misreporting at times when it is most probable.

The conclusion of this analysis, therefore, is that it is not the choice of *signal* which is principally important. Rather, it seems that all potential signals, at annual frequency, are relatively too smoothed for the purpose of validating reported statistics. Reasonably, a higher frequency should be considered, but this will result in the need to drop luminosity from the analysis. Are NO_X emissions a worthy substitute?

5.2 The Global Financial Crisis and Great Recession

We may now conduct a similar analysis for the Great Recession period. Recalling again the discussion of Section 2, the quarterly data set, spanning from 2006 Q1 - 2013 Q3 ($T = 31$) and across all $N = 31$ regions, is utilized.

Again, the first concern is to separate the $N = 31$ regions into smaller groupings in which population structural parameters are feasibly in common. The Monte Carlo experiment summarized by Table 5 indicates that given the longer $T = 31$ length of this dataset, a cross sectional dimension of just $N = 6$ is now likely to lead to correct coverage probabilities. The preliminary estimates of NO_X emissions signal elasticities $\widehat{\beta}$ for this data set, given in Figure A.4, are used to provide an ordering from which to define these groupings.

The results of this objective method of determining groupings is given by Figure 8 (b). Groups 1 and 2 (each $N = 6$) contain high elasticities, while Groups 4 and 5 (each $N = 6$) contain low elasticities. Group 3 ($N = 7$) contains mid-range. The very first item to recognize is that in comparison with the agnostic annual groupings, (Figure 8 (a)), these quarterly regional groupings are transparently more geographically uniform. Specifically, this entirely data-based method of choosing groups puts together regions which are more geographically and economically similar; high elasticity groups are primarily located in the eastern urban provinces, while low elasticity groups are primarily in the western, less urban provinces. Furthermore, that high elasticity groups are located in the east makes sense, as this is the region with relatively more automobile traffic and energy production more generally. Immediately it is clear that quarterly data contains information which is simply not evident in annual, and that this information is economically

Table 7: Estimates: Quarterly sample, 06Q1 - 13Q3.

	Group 1 ($N = 6$)		Group 2 ($N = 6$)		Group 3 ($N = 7$)	
	$S = 1$	$S = 3$	$S = 1$	$S = 3$	$S = 1$	$S = 3$
ϕ_N	-0.02	-1e-4	-0.27	0.99	-1.69*	1e-3
	(-2.64,1.20)		(-0.63,1.27)		(-3.91,-3e-3)	
ϕ_E	0.95	0.34	-0.30	0.01	0.99	0.09
	(-1.13,4.09)		(-0.75,1.21)		(-0.02,2.50)	
ϕ_C	0.78	0.67	-0.51	5e-4	-0.59	0.91
	(-0.87,2.40)		(-0.86,1.00)		(-0.99,1.17)	

	Group 4 ($N = 6$)		Group 5 ($N = 6$)		Pooled ($N = 31$)	
	$S = 1$	$S = 3$	$S = 1$	$S = 3$	$S = 1$	$S = 3$
ϕ_N	-1.61	0	-1.40	0.01	-1.58*	0
	(-3.98,1.82)		(-7.03,1.74)		(-3.11,-0.11)	
ϕ_E	1.03	0.41	6.95	0.01	0.03	1
	(-1.20,2.98)		(-9.65,11.67)		(-1.45,1.59)	
ϕ_C	0.41	0.59	1.08*	0.97	-0.33	0
	(-0.19,1.64)		(0.68,0.86)		(-0.75,1.30)	

Notes: * Significance at 95% confidence level (confidence interval). Groupings are described by Figure 8 (b). N: NO_X emissions. E: Electricity generation. C: Cement production.

sensible.

Quarterly data on luminosity and freight is not available. Table 7 therefore presents signal-independent estimates for the remaining three signals ($S = 1$ columns). Estimates for the structural parameters θ are in the Appendix, Tables A.5-A.6. Once again, confidence intervals tend to be too large to attribute statistical significance to the most important parameters, ϕ and β. Furthermore, group estimates (same table, $S = 3$ columns) again indicate a drowning-out of reported output, as evidenced by their summing to nearly 1 in all cases. This again indicates that signal-by-signal analysis is useful. However, these multiple-signal estimates are informative inasmuch as they indicate which signals are marginally preferred in which groups. Whereas luminosity took that distinction in all cases with annual data, NO_X emissions, electricity generation, and cement production are variably the most preferred across areas in this quarterly sample. Yet, it remains imperative to realize that NO_X emissions is the only of these signals which may be externally validated.

Once again, the more important statistic from the perspective of data validation are the error bands arising from each signal. Quarterly error bands utilizing NO_X emissions are depicted in Table 11. The first observation that must be made with respect to these error bands versus all those computed using annual data is their economic sensibility. Specifically, reported output figures rarely escape these error bands over the time sample. For example, these error bands in most cases do not indicate that the relatively high-growth figures reported before the crisis, or low growth figures reported during the crisis, are unreasonable. Inasmuch, these statistics seem to escape some of the problems experienced when using annualized data and any signal. Similar conclusions are made with respect to error bands using either electricity generation or cement production, Figures A.9 and A.10. Much as the actual groupings selected using quarterly data seem to be more reasonable than using annual data, so do these error bands. This suggests that quarterly data inherently contains more useful information for the purpose of model validation, aside from the secondary problem of which signal is utilized.

In order to focus the discussion of which growth figures are statistically and economically validated, attention is once again given to the four major municipalities in the sample. Error bands computed using each NO_X, electricity generation, and cement production individually over the quarterly period are depicted in Table 12. Note that quarterly NO_X emissions seem to validate reported output numbers in both Beijing and Shanghai over the entirety of the sample. Furthermore, they achieve this whereas all annual signals, and quarterly electricity generation and cement production, fail to. Specifically, as in the annual sample, there is a major slowdown in Beijing prior to the crisis period. However, this variability is primarily within the quarterly NO_X error band, while it insensibly fell outside of the band for any annualized signal. Furthermore, this validation at the quarterly frequency can not be achieved using either electricity generation or cement production, which again indicate the economically insensible conclusion that pre-crisis figures are "too high." Inasmuch, quarterly NO_X emissions are in this case uniquely useful for validating the data.

Are NO_X emissions also useful for determining when reported output is *not* validated? First consider the case of Tianjin. NO_X emissions indicate that reported output figures come just above the error band in the period immediately following the NBER dates. This timing – contemporaneous with a period of economic turmoil – would be consistent with the notion of politically motivated (upward) misreporting. However, it is perhaps too strong to conclude that these figures are necessarily false. For example, while that conclusion is corroborated by cement data, it is contradicted by electricity generation. Even though these later signals are potentially "reconciled" to the data, they are therein useful in this auxiliary role to NO_X.

But now, consider the case of Chongqing. In the midst of the downturn in 2008, the "Chongqing model," brought forth by the city's then Communist Party secretary Bo Xilai, was hailed by many as an exemplary economic initiative. It involved heavy investment, particularly in constructing many highly visible and dazzling skyscrapers, and tax incentives for businesses. While these no doubt contributed to GDP, are the monumental annualized figures reported just following the NBER dates – peaking at a 65% annualized growth rate in 2010 – reasonable?

The substantial distinction of these figures from error bands computed using any NO_X, and corroborated by both electricity and cement, suggest *not*, from a purely statistical perspective. Furthermore, the timing of this distinction fits the political convenience criterion necessary to motivate intentional data manipulation. So, is it reasonable to pinpoint the case of Chongqing as an instance of possible data misreporting? *Yes.* Bo Xiliai gained prestige not only from his political heritage, but ability to put up GDP numbers. He was also an outwardly ambitious leader, who feasibly relied on these figures for advancement. But he is also a famously unscrupulous party official, who was deposed from office and sentenced to life in prison for corruption on September 22, 2013. Even before this fall from grace, his political rivals were known to openly criticize reported GDP figures from Chongqing. In the words of his predecessor Wang Yang in 2009, "Some of our GDP figures sure look rosy." (Liu, 2009). Undoubtedly, some aspect of what Yang calls these "rigged" statistics may have been attributable to actual, though unnecessary production. In his words, building an "unnecessary" bridge just to "tear it down." But what the NO_X emissions based analysis here suggests is that reported GDP figures from Chongqing during this period were also too generous, regardless. In other words, they report production which may not have actually taken place.

In sum, when used in tandem with sound economic reasoning and political context, NO_X emissions do seem to be useful for uncovering cases of possible data manipulation

6 Conclusion

Can officially reported output statistics be externally validated using other verifiable signals of economic growth? This paper has presented evidence that data may be either validated, or proven unlikely, on the basis of satellite measured NO_X emissions data. The results indicate that reported output figures over the Great Recession period are corroborated by satellite readings for many locations, including Beijing and Shanghai. However, reported figures for Chongqing are conspicuously *not* supported, which is meaningful insomuch as this location is now known to have been run by a corrupt political official bent on improving GDP statistics. Furthermore, these inferences are corroborated by indices of industrial production.

In comparison with the benchmark remotely sensed signal – nighttime luminosity – NO_X presents itself as a useful alternative when the nature of the economic problem necessitates higher than annual frequency data be utilized. This paper has also presented a generalization of the model presented in the landmark contribution of Henderson et al. (2012), in which all structural parameters are globally identifiable, and systematic politically motivated misreporting is allowed for without being assumed. It has also developed a statistical framework which allows for the bootstrapping of confidence intervals with correct coverage probability, and error bands for the ultimate combined measure; these error bands are central to discerning when a given reported output statistic is statistically improbable. Yet, as demonstrated in the case of Chongqing and Bo Xilai, only with sound economic reasoning can the stronger conclusion that manipulation is likely be supported. Finally, this model and methodology are generally applicable to any geographical location or time period, as are NO_X emissions.

In conclusion, while the application of the signal, model, and methodology in this paper is of genuine interest in its own right, these tools also present themselves as more broadly useful in a number of economic applications around the globe.

References

Akimoto, H., T. Ohara, J. ichi Kurokawa, and N. Horii (2006). Verification of energy consumption in china during 1996–2003 by using satellite observational data. *Atmospheric Environment 40*(40), 7663 – 7667.

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 1121–1159.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Batson, A. (2010, December 6). Chinese leader called data 'man-made'. Wall Street Journal.

Berenzin, E., I. Konovalov, P. Ciais, A. Richter, S. Tao, G. Janssens-Maenhout, M. Beekmann, and E. D. Schulze (2013). Multiannual changes of CO2 emissions in china: Indirect estimates derived from satellite measurements of tropospheric NO2 columns. *Atmospheric Chemistry and Physics 13*, 9415–9438.

Chang, C., K. Chen, D. F. Waggoner, and T. Zha (2015). Trends and cycles in China's macroeconomy. *Working Paper No. 21244*.

Chen, X. and W. Nordhaus (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Scientists 108*(21), 8589–8594.

Chen, Y., H. Li, and L.-A. Zhou (2005). Relative performance evaluation and the turnover of provincial leaders in China. *Economics Letters 88*(3), 421–425.

Chow, G. (2006). Are Chinese official statistics reliable? *CESifo Economic Studies 52*(2), 396–414.

Fernald, J., I. Malkin, and M. Spiegel (2013). On the reliability of chinese output figures. *Federal Reserve Bank of San Francisco Economic Letter 8*, 1–5.

Fernald, J. G., E. Hsu, and M. M. Spiegel (2015). Is China fudging its figures? Evidence from trading partner data.

Fernald, J. G., M. M. Spiegel, and E. T. Swanson (2014). Monetary policy effectiveness in China: Evidence from a FAVAR model. *Journal of International Money and Finance 49, Part A*, 83 – 103. The Pacific Rim Economies and the Evolution of the International Monetary Architecture.

Fogli, A., E. Hill, and F. Perri (2012). The geography of the Great Recession. In *NBER International Seminar on Macroeconomics 2012*, pp. 305–331. University of Chicago Press.

Ghanem, D. and J. Zhang (2014). 'Effortless Perfection:' Do Chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management 68*(2), 203–25.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both "N" and "T" are large. *Econometrica 70*(4), 1639–1657.

Henderson, V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica*, 1371–1395.

Holz, C. A. (2014). The quality of China's GDP statistics. *China Economic Review 30*, 309–338.

Horowitz, J. L. (2001). The bootstrap. *Handbook of Econometrics 5*, 3159–28.

Itahashi, S., I. Uno, H. Irie, J.-I. Kurokawa, and T. Ohara (2014). Regional modeling of tropospheric NO2 vertical column density over East Asia during the period 2000-2010: Comparison with multisatellite observations. *Atmospheric Chemistry and Physics 14*(7), 3623–3635.

Koch-Weser, I. N. (2013). The reliability of China's economic data: An analysis of national output. *US-China Economic and Security Review Commission Staff Research Project*, 1–32.

Koech, J. and J. Wang (2012). China's slowdown may be worse than official data suggest. *Federal Reserve Bank of Dallas Economic Letter 7*(8), 1–4.

Lin, B. and C. Liu (2016). Why is electricity consumption inconsistent with economic growth in China? *Energy Policy 88*, 310–16.

Lin, J.-T. and M. B. McElroy (2011). Detection from space of a reduction in anthropogenic emissions of nitrogen oxides during the Chinese economic downturn. *Atmospheric Chemistry and Physics 11*(15), 8171–8188.

Liu, M. (2009, September 25). The competition within China's single party. Newsweek.

Liu, Z., D. Guan, W. Wei, S. J. Davis, P. Ciais, J. Bai, S. Peng, Q. Zhang, K. Hubacek, G. Marland, et al. (2015). Reduced carbon emission estimates from fossil fuel combustion and cement production in China. *Nature 524*(7565), 335–338.

Nakamura, E., J. Steinsson, and M. Liu (2016). Are Chinese growth and inflation too smooth? Evidence from Engel curves. *American Economics Journal: Macroeconomics* (Forthcoming).

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 1417–1426.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of DSGE models. *Journal of Monetary Economics 54*(7), 2048–2064.

Rawski, T. G. (2001). What is happening to China's GDP statistics? *China Economic Review 12*(4), 347 – 354.

Rothenberg, T. (1971). Identification in parametric models. *Econometrica 39*(3), 577–91.

Sargent, T. J. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy 97*, 251–287.

Sinton, J. E. (2002). Accuracy and reliability of China's energy statistics. *China Economic Review 12*, 373–83.

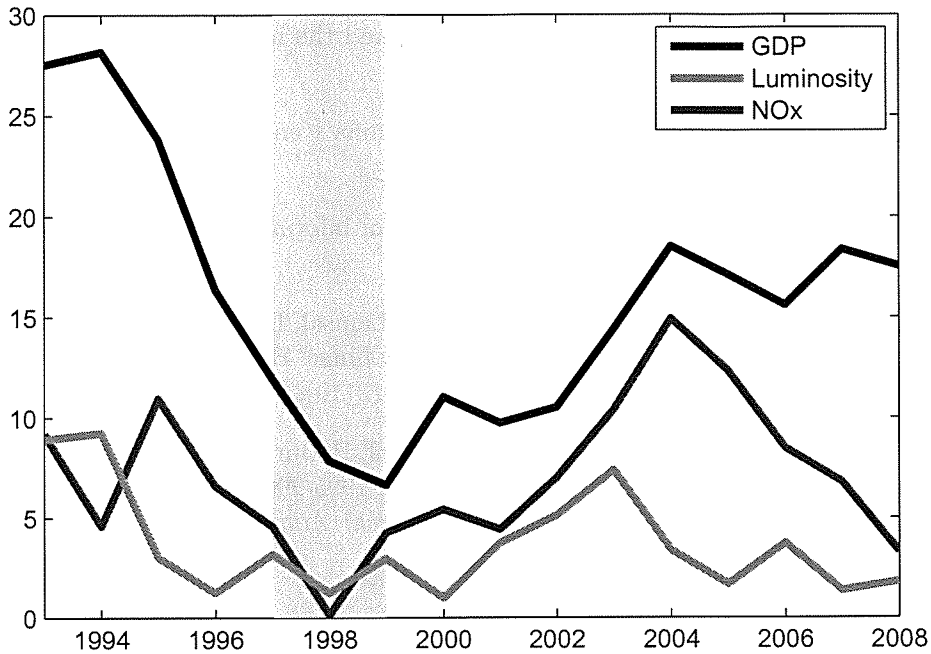
Wallace, J. L. (2014). Juking the stats? Authoritarian information problems in China. *British Journal of Political Science*, 1–19.

Wang, S., Q. Zhang, D. Streets, K. He, R. Martin, L. Lamsal, D. Chen, Y. Lei, and Z. Lu (2012). Growth in NOx emissions from power plants in China: Bottom-up estimates and satellite observations. *Atmospheric Chemistry and Physics 12*(10), 4429–4447.

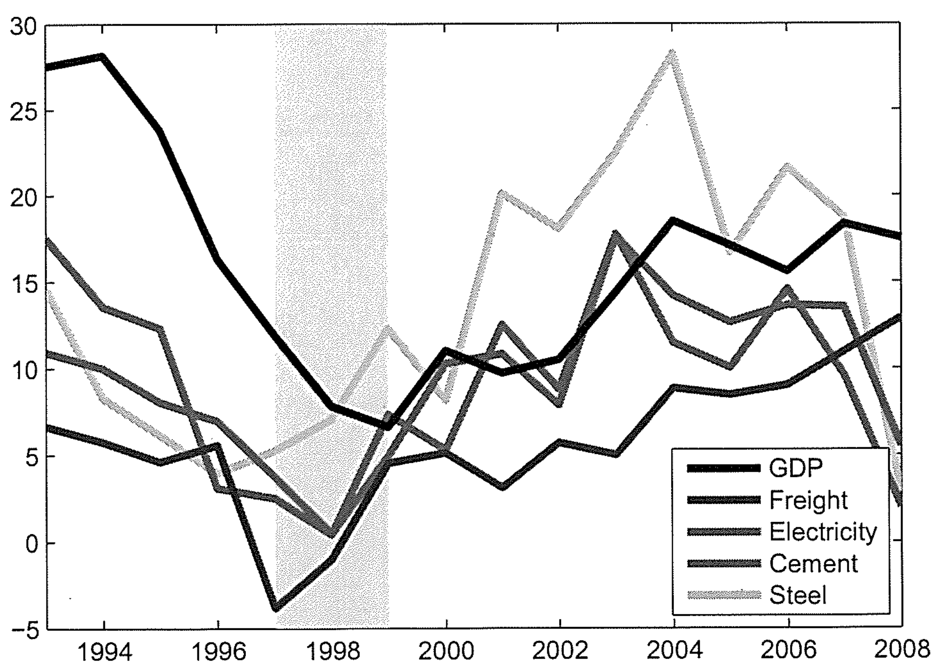
Zhang, Q., D. G. Streets, K. He, Y. Wang, A. Richter, J. P. Burrows, I. Uno, C. J. Jang, D. Chen, Z. Yao, et al. (2007). NOx emission trends for China, 1995-2004: The view from the ground and the view from space. *Journal of Geophysical Research: Atmospheres (1984–2012) 112*(D22).

Figure 4: Annual % change, China: 1993-2008.

(a) Satellite-measured signals.

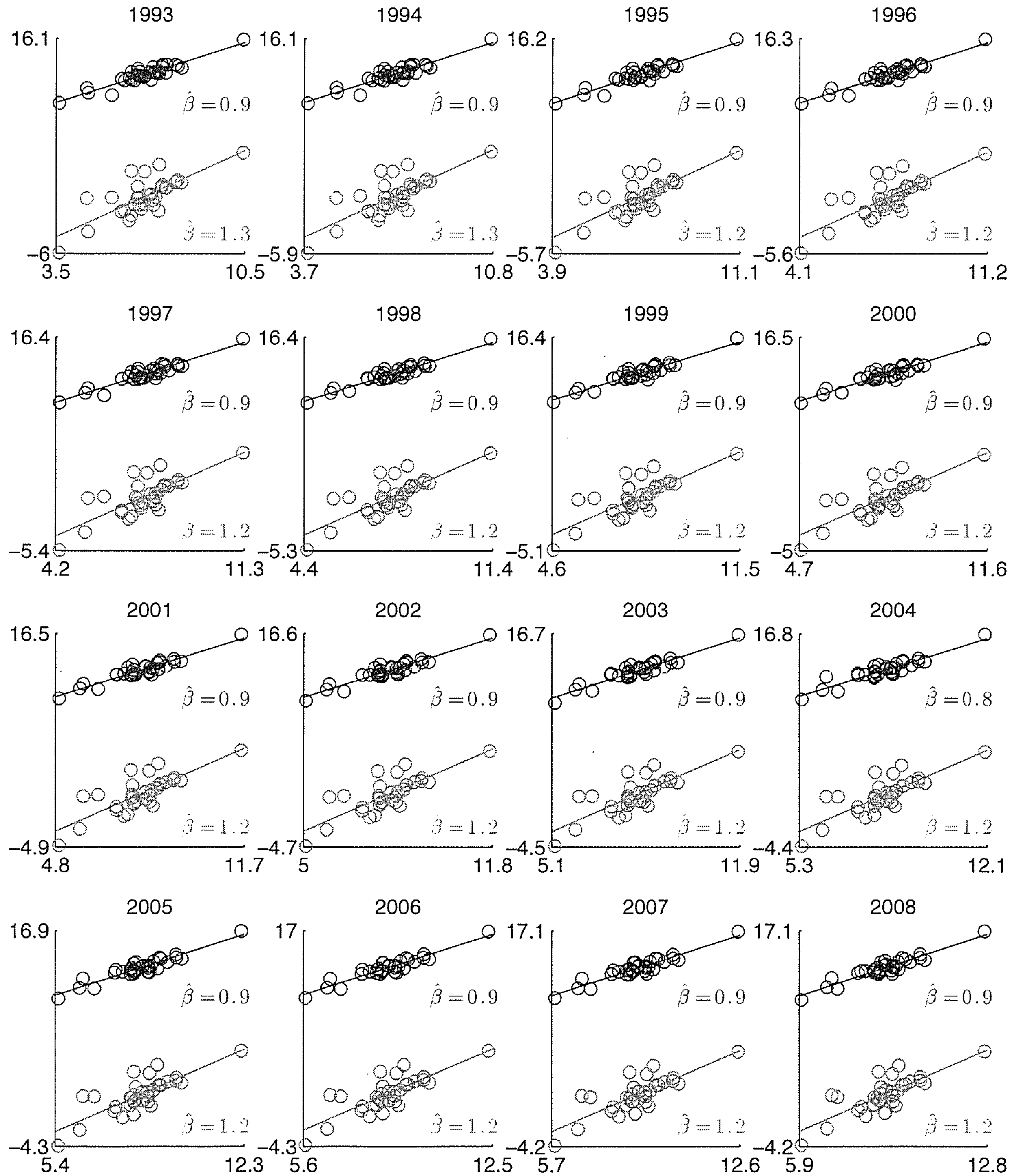


(b) Reported signals.



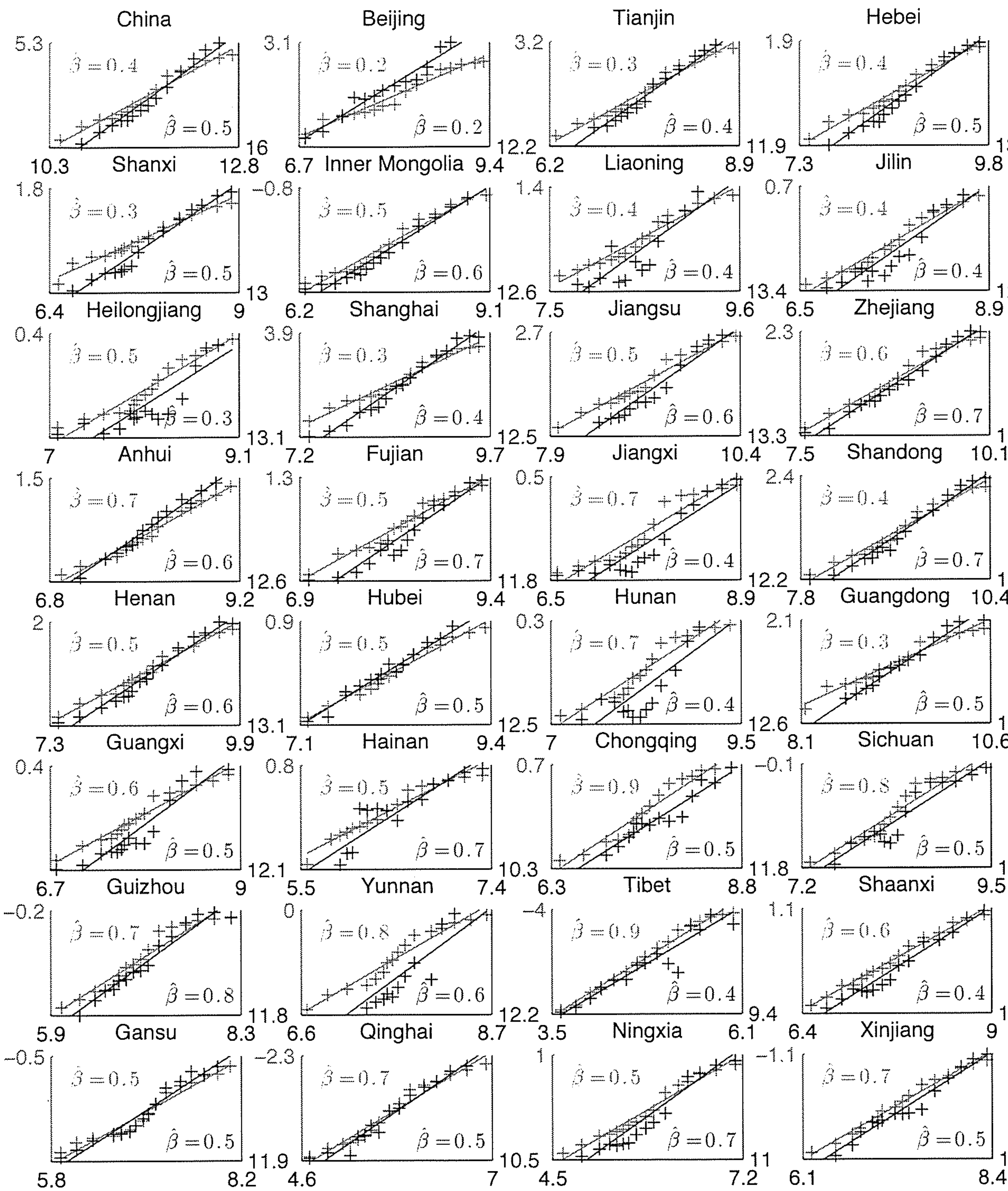
Notes: Shading: Asian Financial Crisis.

Figure 5: Elasticities by-year.



Notes: Pink: Luminosity. Blue: NO_X. Across $N = 31$ regions in each pane. Estimates are for regression model $\ln \text{Signal}_n = \beta \ln \text{Output}_n + \varepsilon_n$ for each t.

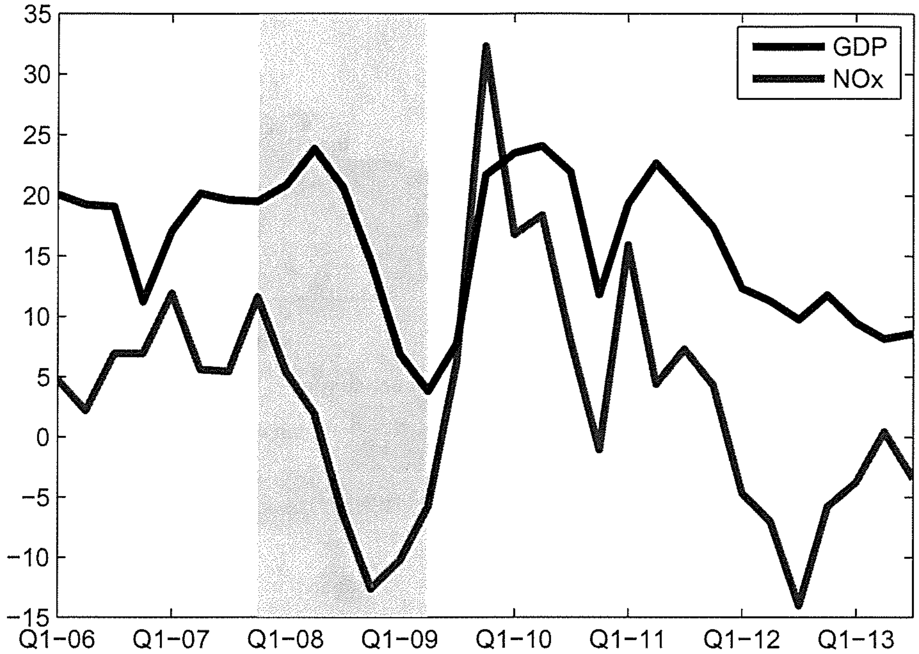
Figure 6: Annual elasticities by-region (1993-2008).



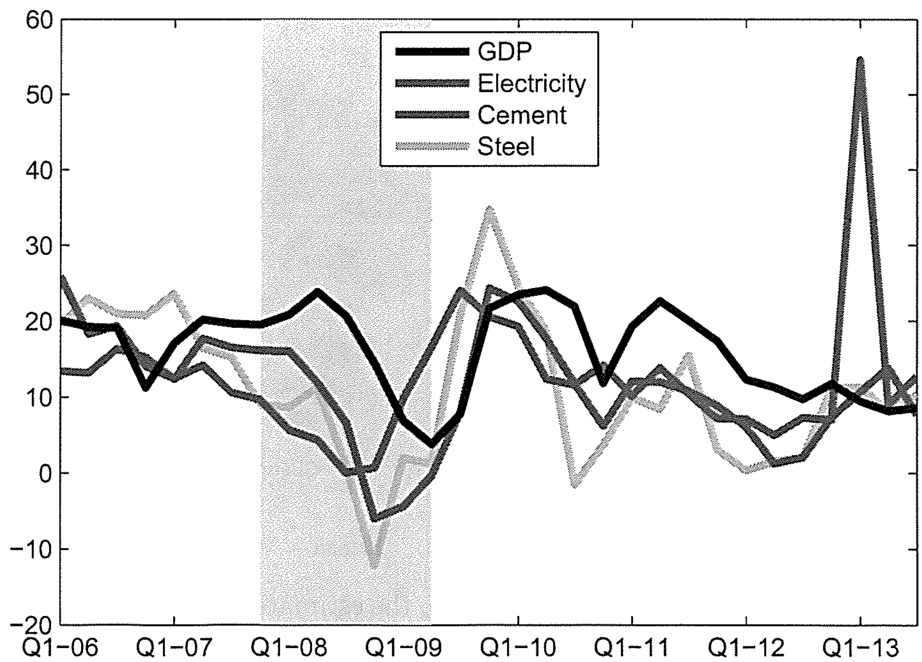
Notes: Pink: Luminosity. Blue: NO_X. Across $T = 16$ quarters in each pane. Estimates are for regression model $\ln \text{Signal}_t = \beta \ln \text{Output}_t + \varepsilon_t$ for each n.

Figure 7: Annualized quarterly % change, China: 2006 Q1 - 2013 Q3.

(a) Satellite-measured NO_X.



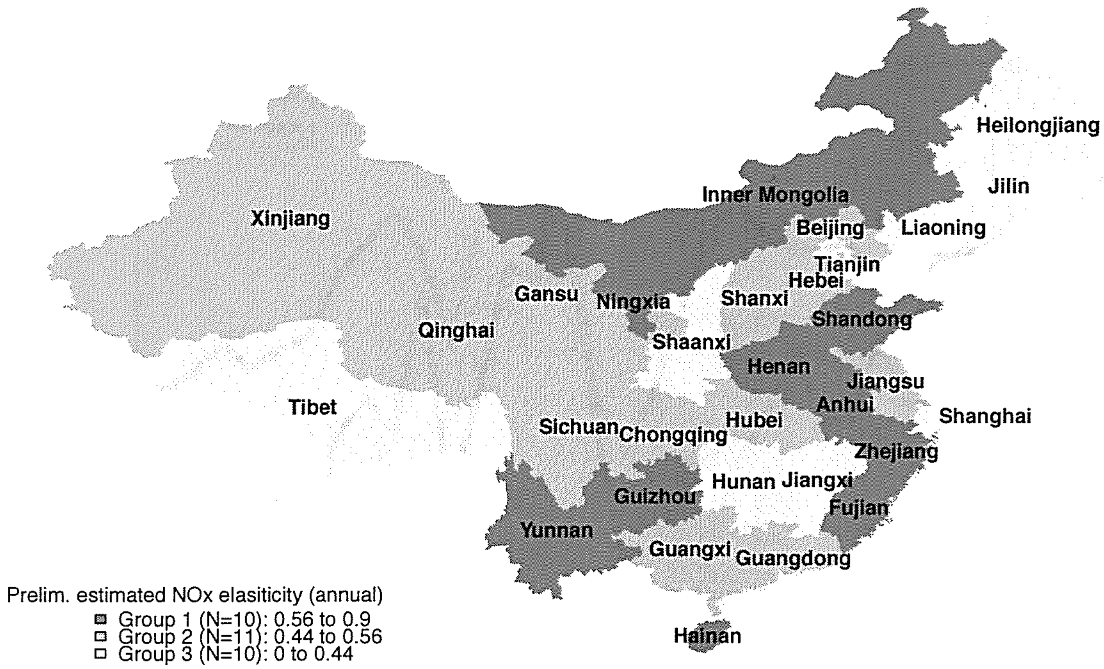
(b) Reported signals.



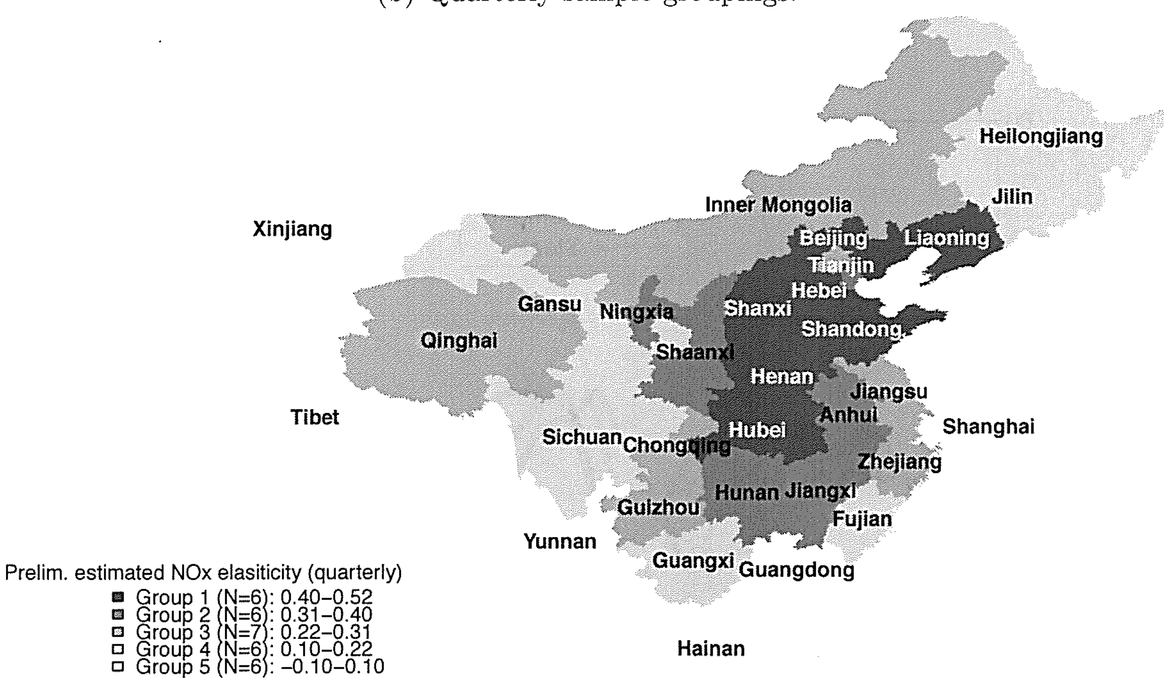
Notes: Shading: NBER recession dates.

Figure 8: Estimation groupings.

(a) Annual sample groupings.

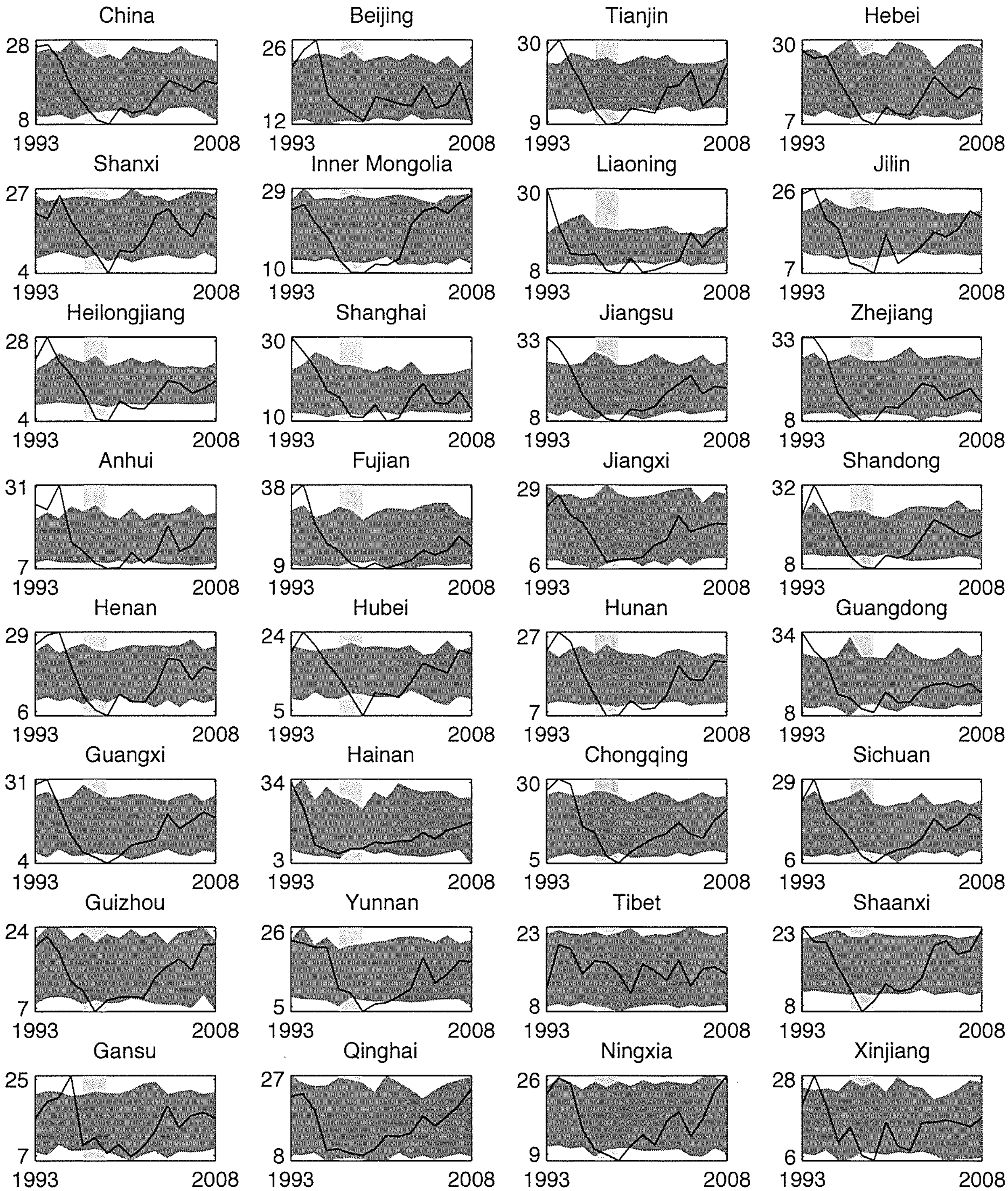


(b) Quarterly sample groupings.



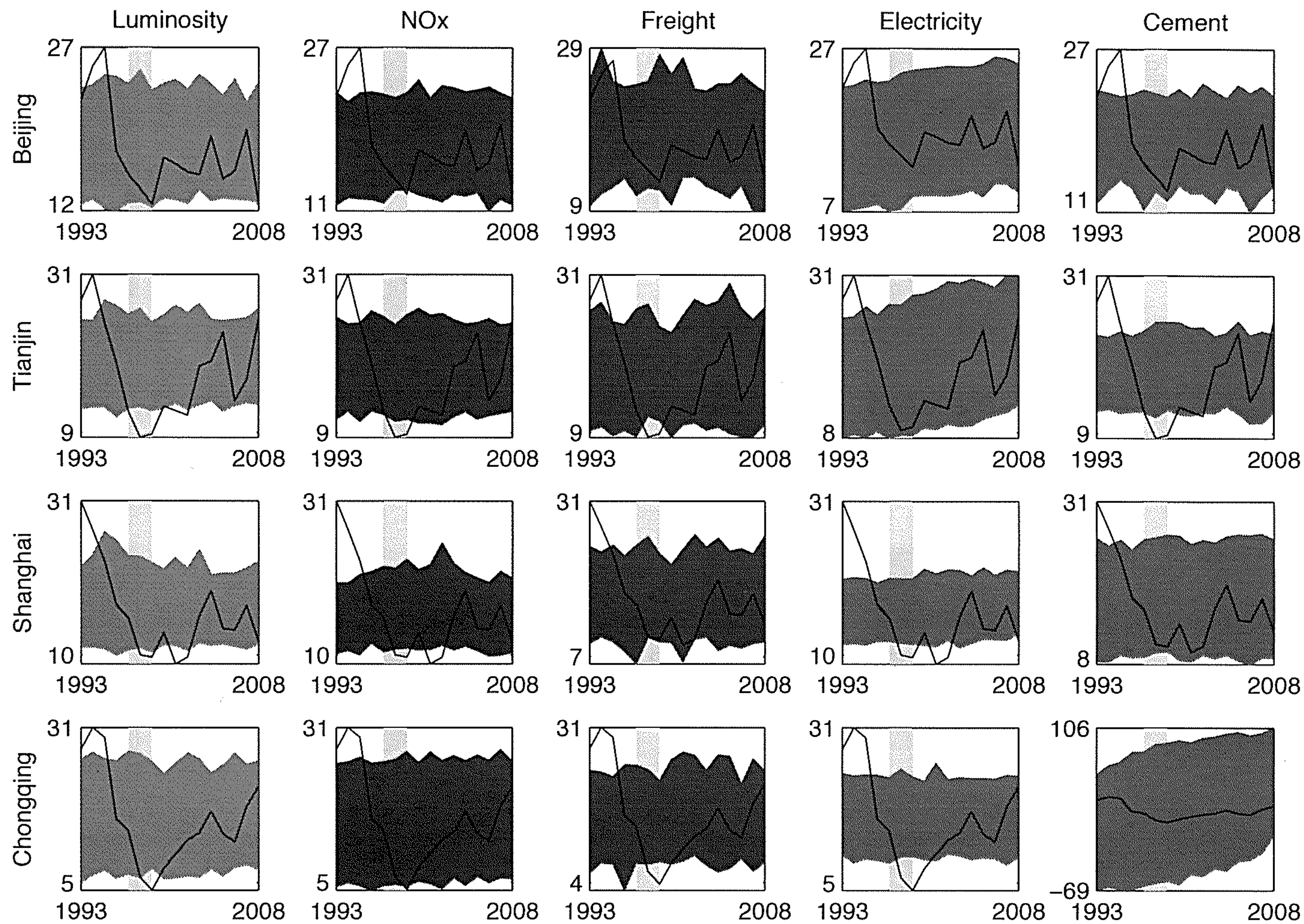
Notes: "Preliminary estimated NO_X elasticity" refers to the estimates $\widehat{\beta}$ computed using NO_X emissions data in Figure 6 for the annual sample, and Figure A.4 for the quarterly sample.

Figure 9: Annual luminosity 95% error bands.



Notes: Black line: Annual reported % change, regional output. Gray shading: 1997-1999 Asian Financial Crisis period. Pink shading: Error bands. Error bands are computed by bootstrap using each respective Group 1-3 estimates. China computed using "Pooled" estimates.

Figure 10: Annual 95% error bands: Municipalities by signals.

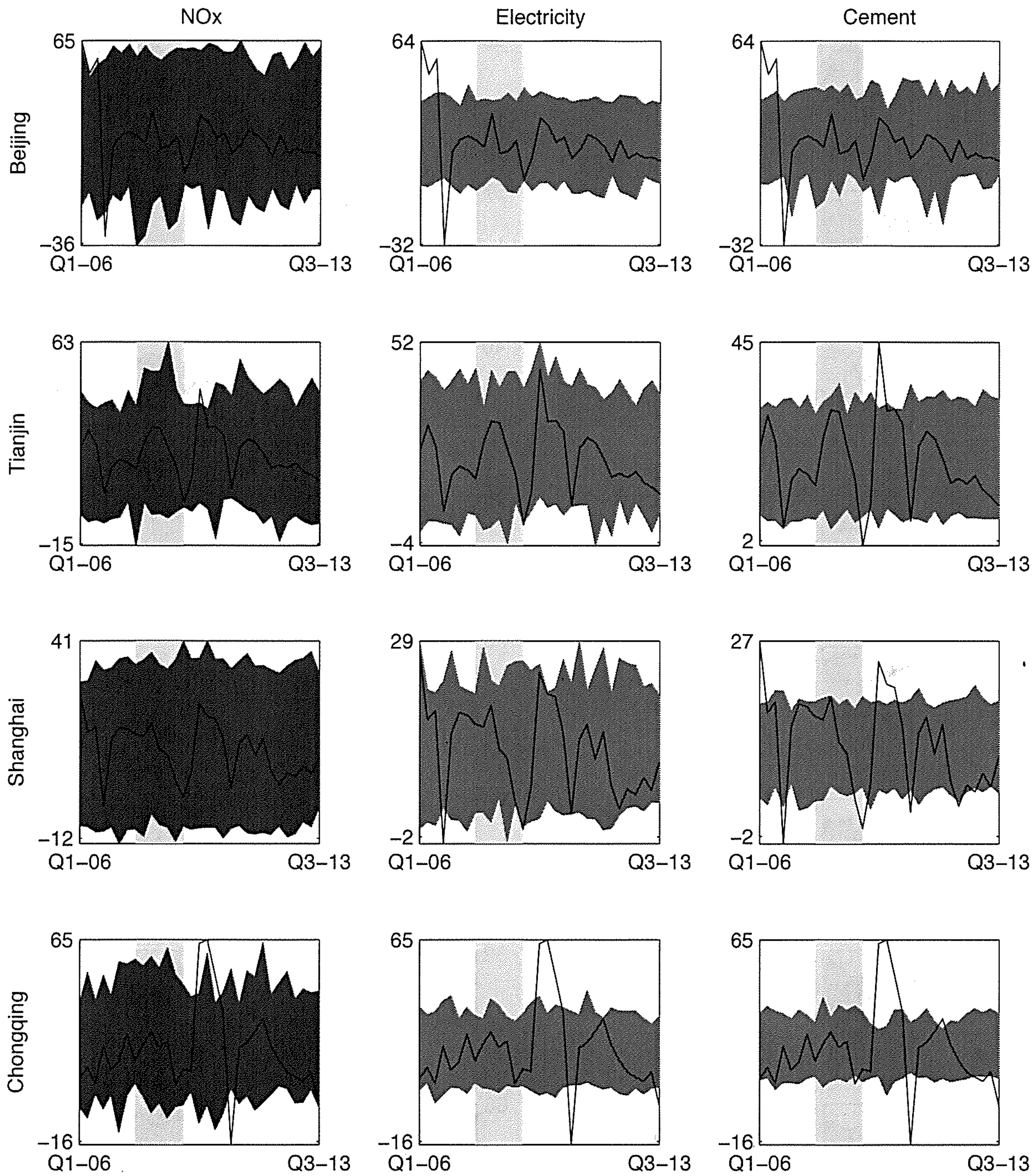


Notes: Black line: Annual reported % change, regional output. Gray shading: 1997-1999 Asian Financial Crisis period. Colored shading: Error bands. Error bands are computed by bootstrap using each respective Group 1-3 estimates. Beijing, Tianjin, and Shanghai are in Annual Group 1. Chonqing is in Group 2.

Figure 11: Quarterly NO_x 95% error bands.

Notes: Black line: Annualized quarterly reported % change, regional output. Gray shading: 2007-2009 Global Financial Crisis period. Pink shading: Error bands. Error bands are computed by bootstrap using each respective Group 1-5 estimates. China computed using "Pooled" estimates.

Figure 12: Quarterly 95% error bands: Municipalities by signals.



Notes: Black line: Annualized quarterly reported % change, regional output. Gray shading: 2007-2009 Global Financial Crisis period. Colored shading: Error bands. Error bands are computed by bootstrap using each respective Group 1-5 estimates. China computed using "Pooled" estimates. Shanghai is in Quarterly Group 1. Beijing and Chongqing are in Group 3. Tianjin is in Group 4.