Macroeconomic Dynamics, **23**, 2019, 3327–3354. Printed in the United States of America.
doi:10.1017/S1365100518000056

VALIDATING CHINA'S OUTPUT DATA USING SATELLITE OBSERVATIONS

Stephen D. Morris
Bowdoin College

Junjie Zhang
Duke Kunshan University
Duke University

Can officially reported output figures be externally validated? This paper presents a dynamic panel framework for assessing statistics using verifiable signals of economic activity. In this context, satellite readings of nitrogen dioxide, a byproduct of combustion, are forwarded. The problem of validating China's reported gross domestic product at the sub-national level during two recent downturns is considered. During the Great Recession period, reported figures are validated for some regions, but not others, including specifically those known to be inaccurate.

Keywords: China, GDP, Nitrogen Dioxide, Night Lights

1. INTRODUCTION

Can officially reported output figures be externally validated? This question is pertinent to data from China, which have long been cast under suspicion. While opaque methods in data collection represent one cause for concern,[1] another is systematic 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 in GDP. This figure appears to be an aberration not only

The Online Appendix to this paper is available at `https://goo.gl/E9G5b8`. Data sets are freely available at `https://goo.gl/v2Dk4S` and replication materials at `https://goo.gl/hVrGWD`. We thank Qiang Zhang at Tsinghua University for providing the data. We 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, the 2014 Workshop in Macroeconomic Research at Liberal Arts Colleges at Wesleyan University, and the 2016 conference "The Environment: Issues and Strategies in China & The U.S" at UC Santa Barbara, as well as many individuals, for productive discussions. Finally, we thank the Associate Editor and two anonymous referees for useful comments. Errors are our own. Address correspondence to: Stephen D. Morris, Bowdoin College, Department of Economics, 9700 College Station, Brunswick, ME 04011, USA; e-mail: smorris@bowdoin.edu.

 1365-1005/18

regionally, but also in the context of sharper contemporaneous declines in energy consumption [Rawski (2001)]. Similar concerns arose during the Great Recession period of 2007–2009, an era from which inflation and consumption data are suspected [Nakamura et al. (2016)]. But pairwise incompatibilities of reported GDP with other conventional indicators of growth, such as electricity generation, were evidently more muted during this later episode. In fact, both statistics indicate a healthy rise in economic activity over the NBER recession dates, 2007 Q4–2009 Q2 [Figure 1 (a)].[2] Perhaps, due to this redundancy of information, economists have in this case come to more mixed conclusions as to the veracity of reported statistics; Fernald et al. (2013) argue that industrial production indices including electricity generation corroborate reported data, while the opposite case had previously been made by Koech and Wang (2012).

But perhaps the true problem is that electricity generation is reported by the very same officials who report output. This paper makes three contributions in this arena. First, independently reported international satellite readings of *nitrogen dioxide* (NO_2), a byproduct of combustion, are forwarded as a useful signal from which to construct a proxy for economic growth. This data set is freely available, on a dense geographic mesh of the Earth's habitable surface, for a time sample of the past 20 years. We suggest that this data set is a useful companion to another truly unadulterated signal of growth, night lights [Henderson et al. (2012)]. Inter-quarterly business cycle dynamics are of interest, but night lights data are available only annually; NO_2 is reliable up to monthly frequency, and available up to daily. In contrast with electricity generation, NO_2 suggests a significant downturn in economic activity in China during the Great Recession period [Figure 1 (b)].

Second, an econometric framework for constructing combined measures of economic growth using any proxy is presented. In contrast with existing approaches, this dynamic panel data (DPD) framework exploits not only the cross-sectional, but also time series dimension of the data. This enables the analysis to focus on particular sub-national regions over the business cycle, which is critical since variability in economic activity across these regions is substantial [Figures 1(c) and (d)].[3] Moreover, as we discuss, it is the provincial level of the political hierarchy at which misreporting may actually occur in practice, focusing on the national level is prone to obscuring these dynamics. Central to this analysis is the knowledge of the elasticity of a signal with respect to income, as well as other structural parameters. We show that the identification and estimation of these parameters is subject to a number of previously undocumented econometric challenges. These include nuisance parameters which hinder the interpretability of reduced form vector autoregressive (VAR) estimates, biases, and under-coverage of asymptotic confidence intervals in small samples.

As part of this analysis, we also show how to compute the sampling distribution of the weighting on the proxy in the combined measure. Our third theoretical contribution is to show that in fact, neither the weighting on the proxy created using NO_2, nor night lights, nor any of a number of indices of industrial pro-

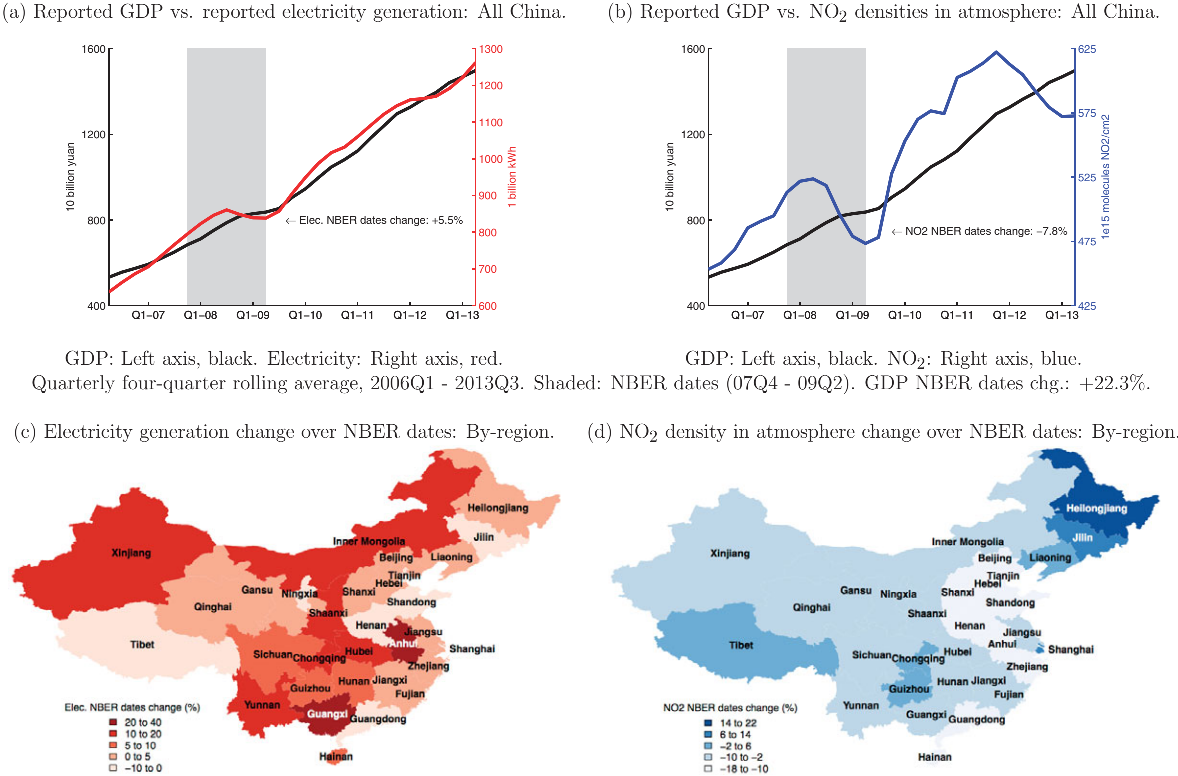

FIGURE 1. Reported GDP and signals of economic growth: China, 2006–2013. (a) Reported GDP vs. reported electricity generation: All China. (b) Reported GDP vs. NO_2 densities in atmosphere: All China. (c) Electricity generation change over NBER dates: By-region. (d) NO_2 density in atmosphere change over NBER dates: By-region.

duction, is statistically significant. This calls into question the interpretability of point estimates of combined measures using any signal.[4] Instead, this paper shows how to compute the sampling distribution of the combined measure by bootstrap, and notes that these confidence bands are nonetheless useful for *validating* reported GDP figures. This is therefore the extent of the empirical question one may reasonably hope to answer, and the one which we pursue in this paper.

In the results, we show that if one attempts to consider aggregate national data, then one fails to reject that officially reported GDP statistics during the Great Recession period are validated. But at the sub-national level, statistics from certain provinces, autonomous regions, and municipalities are not validated. A subset of these entirely satellite-identified areas corresponds to those now known to have embellished data, due to later revelations from a 2015 corruption probe. In other words, the methodology is also proven to work as intended.

2. DATA SET

Statistics which are merely indicative of production 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 the signal with respect to income. We now consider what make useful signals, the panel of interest in this paper, and how elasticities may vary or remain fixed across time and space.

2.1. Signals

Why might officially reported indices of industrial production such as electricity generation make problematic signals? Two assumptions are always necessary to identify the elasticity of growth with respect to the signal: (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 these statistics are reported by the same entities reporting output, casting a shadow over their credibility.[5] 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)].[6]

Given these concerns, it is necessary to seek out signals which are guaranteed to be free of distortion. The benchmark in the realm of externally verifiable growth measurement is night-time luminosity, or “night lights.” This signal is not only naturally indicative of energy consumption, but measured by orbiting international satellite instruments, and publicly available. Henderson et al. (2012) pioneered the application of this data set to formally producing *combined measures* of economic growth—optimal weighted averages of proxy and reported output—in areas where data collection is otherwise challenging. This data set has proven useful in many contexts, particularly in development economics. However, the DMSP–OLS luminosity data set currently only ranges from 1992 to 2013, an annual time dimension of just $T = 22$. This is limiting of time-domain business cycle analysis.

This paper suggests satellite readings of tropospheric NO_2 densities as another useful signal of economic activity. NO_2 is a byproduct of anthropogenic sources primarily including combustion, and therefore directly indicative of economic activity. Natural sources of NO_2 including biomass burning, soil, and lightning comprise errors in the interpretability of NO_2 as a signal which, later, we model explicitly. The satellite measurements we consider take the form of vertical column densities (VCDs or "columns") which are geographically gridded measurements (up to a few square kilometers) of trace gas concentrations. VCDs are inferred by satellites as a function of the difference between sunlight scattered in the atmosphere versus that reflected by the Earth. High spatial resolution and sampling frequency (up to daily) enables one to assemble a longitudinal data set which like luminosity is guaranteed to be free of political influence, and indicative of growth.[7] But unlike luminosity, NO_2 data is reliable at least up to a monthly basis, allowing one to consider business-cycle fluctuations at the regional level. The satellite VCD readings of NO_2 utilized in this paper are based on quarterly mean tropospheric NO_2 from Ozone Monitoring Instrument, a UV/Vis nadir spectrometer onboard NASA's EOS-Aura satellite.[8] These data are publicly, freely available online at sources including `http://www.temis.nl/`.

There is precedent for using NO_2 columns to infer otherwise unobservable data from China. *Nitrogen oxides* (NO_x) is a generic umbrella term encapsulating both nitric oxides (NO) and NO_2. Anthropogenic emissions of NO_x are of keen interest to scientists, as they affect the formation of ozone, and may have harmful environmental implications including smog and acid rain. Such emissions have surged in China over the past decades coinciding with increased economic growth. For locations elsewhere in the world, 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 part of local figureheads' annual performance reviews, 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 of columns [Wang et al. (2012)].[9] NO_x emissions data are more directly indicative of human activity than NO_2 densities in that nonanthropogenic contributors are sorted out. However, these data are less readily available to economists, since it first must be estimated using a scientific model.

Finally, among all tropospheric species, NO_2 is arguably the easiest to measure. The short atmospheric life of NO_2 means that densities are closely correlated to emissions, and therefore may be used to accurately measure them independent of wind or other obfuscation.[10]

TABLE 1. Data set

$N = 31$ 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	tons	✓	
	NO_2 columns	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, and Chongqing) and autonomous regions (Guangxi, Inner Mongolia, Tibet, Ningxia, and Xinjiang). *Tibet missing. **Tibet, Hainan, Ningxia missing.

2.2. Sample

Two separate panels are assembled, as described by Table 1. Both panels correspond to the same set of $N = 31$ provinces, municipalities, and autonomous regions previously depicted in Figures 1(c) and (d). But the panels differ in the time T dimension. The time periods selected purposefully subsume two economically challenging periods in China, during which systematic manipulation of data at the sub-national level may have been more likely.

The first time period, at the annual frequency from 1993 to 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 usefulness of purely anthropogenic NO_x emissions data versus night-time luminosity. In addition, several reported indices of industrial production including electricity generation are included.[11]

The second period, at the quarterly frequency from 2006 Q1–2013 Q3, includes the Great Recession of 2007–2009. Both luminosity and freight traffic must be dropped, as neither is available at higher than annual frequency. In this time sample, we revert to simply using raw NO_2 column data, rather than NO_x emissions. Recall, this data naturally contains errors due to nonanthropogenic sources of NO_2 densities. However, there is good reason to believe this data in itself is a useful indicator of human activities specifically.[12] Moreover, it is readily available to economists, unlike NO_x emissions, so we wish to understand if it is equally informative.

2.3. Elasticities

Utilizing any signal to construct a proxy for growth relies on identifying the elasticity of the signal with respect to income, β. If this signal is constant across

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

regions $n = 1, \ldots, N$ and time periods $t = 1, \ldots, T$, then the following equality will hold with independently and identically distributed error ε_{nt}:

$$\%\Delta\text{Signal}_{nt} = \beta \times \%\Delta\text{Output}_{nt} + \varepsilon_{nt}. \tag{1}$$

Estimating the elasticity as a fixed parameter with cross-sectional data would require that β is fixed across n. But this does not seem to be true in the data.[13] Intuitively, each area is economically distinct, and variety technologies and utilized inputs result in distinct elasticities of each signal with respect to output; one need only consider the distinction in technologies utilized in the largely rural western provinces of China versus the largely urban centers in the east.[14] One may further corroborate this result by analyzing the time dimension cross-correlations for each individual area.[15] In the annual sample, these correlations fluctuate across regions; for instance, the correlation of GDP with luminosity is 0.81 in Hainan but 0 in Sichuan.[16] So, in both samples, there is ample evidence that signal elasticities are not constant across regions in China, a result which holds for all signals. In this sense, β is *not* constant across n, and exploiting large N results to estimate (1) is unlikely to result in a consistent estimator.

Given these observations, the perspective of this paper is that cross-sectional areas should be small enough so that elasticities are confidently fixed across them, and then time-series variation should be introduced to supplement the empirical analysis. Estimating the elasticity as a fixed parameter with time series data would require that β is fixed across t. First, it must be addressed that this may not strictly be so in the longer run, when technical change could be substantial. For example, we might expect to observe the elasticity of NO_x to decay as a result of substitution toward less pollutant i.e. greener technologies over a long enough horizon. In the time samples utilized in this analysis, 8 and 16 years, there appears to be evidence that any such trend is small.[17] Nonetheless, it remains critical to investigate this point further, and establish that if such trends exist, they are sufficiently slow moving for NO_x to be useful as a signal. In the following sections, while we assume the signal–output relationship is constant, in Section 7, we show how to directly verify this assumption using two statistical tests.

3. MODEL

The question of computing externally verifiable growth statistics for China and other areas has been considered from the cross-sectional and time-series dimensions independently. Henderson et al. (2012) construct combined measures of economic growth for many global regions (excluding China) at certain points in time using cross-sectional night lights data. Fernald et al. (2015) use trading partner data to assess growth statistics in China at the national level from the time series perspective.[18]

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

We have seen that it may be more useful to consider the case of China from the time series perspective, as elasticities may differ substantially across regions. But why may it be important to consider the question of economic growth in China at specifically the *sub-national* level over time? 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 cited through Wikileaks cables as using electricity consumption, volume of rail cargo, and the amount of loans disbursed as signals from which to infer his measure [e.g. Batson (2010)]. While this quotation is widely known, it has a more subtle implication that is often overlooked. Specifically, the fact that Party elite would be forced to rely on such a proxy offers perspective on an aspect of Chinese political economy. The national government of China naturally desires accurate statistics for the purpose of policy making, and theories of manipulation at this vantage are usually refuted [Chow (2006)].[19] 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 containing GDP statistics from their district [Chen et al. (2005)]. As evidence of the competing objectives of local and national officials, beginning in 1998, China's National Bureau of Statistics began to bypass some provincial governments in data accumulation. In sum, it is the local-level change in signals which is relevant for proxy construction, as this is the scope at which output figures may be inaccurate.

This section offers refinements in the arena of combined measure construction pioneered by Henderson et al. (2012). The primary objective will be to extend this framework from a cross-sectional, to dynamic panel perspective, where large-T results will be applicable.

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 observable.

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[20]

$$y^*_{nt} = \rho_y y^*_{nt-1} + \varepsilon^y_{nt}. \tag{2}$$

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

The following standard regularity conditions are also assumed: $0 < |\rho_y| < 1$, $\varepsilon^y_{nt} \sim \text{IWN}(0, \sigma^2_y)$ (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.

Equation (2), which imposes dynamics on the system, is unique to the model specification of this paper, versus Henderson et al. (2012)'s setup. However, the next three equations are largely consistent. Say that that the latent or otherwise unobservable 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

$$s^*_{nt} = \beta y^*_{nt}, \tag{3}$$

$\beta \neq 0$ 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. NO_2 column data is subject to meteorological and other nonanthropogenic error. Henceforth, this paper adopts the notation that variables without-* (s_{nt}) represent the observable datum corresponding to each with-* unobservable counterpart (s^*_{nt}).

$$s_{nt} = s^*_{nt} + \varepsilon^s_{nt}, \tag{4}$$

$\varepsilon^s_{nt} \sim \text{IWN}(0, \sigma^2)$ 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 only require that scientific or other exogenous error in measuring the signal is idiosyncratic, and unrelated to business cycle fluctuations.

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

$$y_{nt} = y^*_{nt} + u^*_{nt}. \tag{5}$$

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

$$u^*_{nt} = \rho_u u^*_{nt-1} + \varepsilon^u_{nt}, \tag{6}$$

$0 < |\rho_u| < 1, \varepsilon^u_{nt} \sim \text{IWN}(0, \sigma^2_u)$ 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

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

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. (2012). The purpose of this generalization is both statistical and structural. From a statistical perspective, the basic model (2) and (3), 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.[21] From a structural perspective, since u^*_{nt} is persistent, its process also allows for systematic human-related intervention in data reporting.[22]

3.2. Combined Measure

Equations (2)–(6) encapsulate the model. In total, there are six structural parameters, collected in

$$\underset{6\times 1}{\theta} = (\underbrace{\beta, \sigma}_{\text{Signal}}, \underbrace{\rho_y, \sigma_y}_{\text{Output}}, \underbrace{\rho_u, \sigma_u}_{\text{Error}})'. \tag{7}$$

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) and (4), one may construct a proxy for growth. It is the linear projection

$$\widehat{z}_{nt} = \left(1/\widehat{\beta}\right) s_{nt}. \tag{8}$$

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 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

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

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

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

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

Simply rearranging yields the optimal loading:

$$\widehat{\phi} = \frac{\widehat{\sigma}_u^2}{1-\widehat{\rho}_u^2} \Big/ \left(\frac{\widehat{\sigma}_u^2}{1-\widehat{\rho}_u^2} + \frac{\widehat{\sigma}^2}{\widehat{\beta}^2} \right). \tag{9}$$

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

$$\widehat{x}_{nt} = (1-\widehat{\phi}) y_{nt} + \widehat{\phi}\widehat{z}_{nt}. \tag{10}$$

$\widehat{\phi} \in [0, 1]$ 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.[23] 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 might 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:

$$\underset{S\times 1}{s_{nt}} = \left[s_{nt}(1) \ldots s_{nt}(S)\right]', \qquad \underset{S\times 1}{s_{nt}^*} = \left[s_{nt}^*(1) \ldots s_{nt}^*(S)\right]',$$

$$\underset{S\times 1}{\varepsilon_{nt}^s} = \left[\varepsilon_{nt}^s(1) \ldots \varepsilon_{nt}^s(S)\right]'.$$

Henceforth, parenthetic (i) denotes parameters and variables which are signal i-specific. The system equivalent of the signal measurement equation (4) is again written,

$$s_{nt} = s_{nt}^* + \varepsilon_{nt}^s, \tag{11}$$

but now, s_{nt}, s_{nt}^*, and ε_{nt}^s are vectors. In the case of a lone signal $S = 1$, there are six structural parameters in θ (7). The first two (β, σ) are signal-dependent, while the latter four $\left(\rho_y, \sigma_y, \rho_u, \sigma_u\right)$ 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 among signals. Thus, there are $n_\theta = S + S(S+1)/2 + 4$ total structural parameters for $S \geq 1$ signals.

$$\underset{n_\theta \times 1}{\theta} = (\underbrace{\beta', \sigma'}_{S \text{ Signals}}, \underbrace{\rho_y, \sigma_y}_{\text{Output}}, \underbrace{\rho_u, \sigma_u}_{\text{Error}})'. \tag{12}$$

$$\underset{S\times 1}{\beta} = \left[\beta(1) \ldots \beta(S)\right]' \qquad \underset{S(S+1)\times 1}{\sigma} = \mathrm{vech}(L) \qquad \underset{S\times S}{L} = \mathrm{chol}(\Sigma)$$

$$\underset{S\times S}{\Sigma} \equiv E(\varepsilon_{nt}^{s}\varepsilon_{nt}^{s'})$$

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. Finally, note that in the special case that $S = 1$, then $n_\theta = 6$, and we have the previous lone signal model setup.

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

$$\underset{S\times 1}{\widehat{z}_{nt}} = \mathrm{diag}(\widehat{\beta})^{-1} s_{nt}. \tag{13}$$

Recall, the individual elasticities are nonzero by assumption of the signals' relevance. Therefore, 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}, which is

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

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

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

Simply rearranging yields the optimal loadings:

$$\underset{S\times 1}{\widehat{\phi}} = \left[1_S 1_S' \frac{\widehat{\sigma}_u^2}{1 - \widehat{\rho}_u^2} + \mathrm{diag}(\widehat{\beta})^{-1} \widehat{\Sigma} \mathrm{diag}(\widehat{\beta})^{-1}\right]^{-1} \times 1_S \frac{\widehat{\sigma}_u^2}{1 - \widehat{\rho}_u^2}. \tag{14}$$

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

$\widehat{\phi}'1_S \in [0, 1]$ 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 (9). However, (14) is not equivalent to a vector of (9) 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

$$\widehat{x}_{nt} = (1 - \widehat{\phi}'1_S) y_{nt} + \widehat{\phi}'\widehat{z}_{nt}. \tag{15}$$

4. METHODOLOGY

This section first establishes that the model observables have DPD representation with serially correlated errors. 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 make any attempt to decipher latent growth from a VAR, or other linear reduced form model, subject to misleading conclusions. Finally, this ultimately computationally efficient estimator also expedites the bootstrapping of small T sample bias correction, confidence intervals for the structural parameters and proxy loading, and confidence bands for the combined measure.

The model (2)–(6) contains both observable and unobservable variables, implying that it has state space representation when utilizing any lone signal. Proposition 1, proven in Online Appendix Section C, establishes that this state space also yields parsimonious reduced form representation for the observables given any $S \geq 1$ set of signals.

PROPOSITION 1 (Reduced form representation). *The observables have the representation*

$$Y_{nt} = X_{nt}\Psi + V_{nt} \tag{16}$$

for $X_{nt} = \left(Y'_{nt-1} \otimes I_{S+1}\right) R$ *is* $(S+1) \times (S+2)$, $Y_{nt} = \left[y_{nt}\ s'_{nt}\right]'$ *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*

$$\Psi(\theta; \lambda) = \left[\rho_y\ \rho_u\ \psi'\right]', \tag{17}$$

$$\underset{S\times 1}{\psi}(\theta; \lambda) = (\rho_y - \rho_u)\left[\frac{\lambda_1}{\beta(1)} \cdots \frac{\lambda_S}{\beta(S)}\right]', \tag{18}$$

$$\underset{S\times 1}{\lambda} = \left[\lambda_1 \ldots \lambda_S\right]'; \qquad \lambda' 1_S = 1, \tag{19}$$

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

$$\Omega(\theta; \lambda) = E(V_{nt} V'_{nt}) \tag{20}$$

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

Equation (16) 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.

From a panel perspective, however, this representation is more naturally interpreted as a realization of 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 not even yield a consistent estimator. Second, the fact that $\{V_{nt}\}$ is serially correlated means that estimators which assume otherwise are subject to endogeneity problems. 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. In certain cases, this might make the interpretation 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)], unclear.

While we are ultimately interested in estimation and inference for θ, such an analysis is impossible unless it is first possible to identify Ψ and Ω. Proposition 2, proven in Online 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*

$$\underset{(S+2)\times 1}{\widehat{\Psi}} = \left[\sum_{n=1}^{N} \widehat{Z}_n' \widehat{X}_n\right]^{-1} \sum_{n=1}^{N} \widehat{Z}_n' \widehat{Y}_n, \tag{21}$$

$$\underset{T(S+1)\times 1}{\widehat{Y}_n} = \left[\widehat{Y}_{n3}' \ \ldots \ \widehat{Y}_{nT+2}'\right]', \qquad \underset{T(S+1)\times(S+2)}{\widehat{X}_n} = \left[\widehat{X}_{n3}' \ \ldots \ \widehat{X}_{nT+2}'\right]',$$

$$\underset{T(S+1)\times(S+2)}{\widehat{Z}_n} = \left[\widehat{X}_{n2}' \ \ldots \ \widehat{X}_{nT+1}'\right]', \qquad \widehat{X}_{nt} = (\widehat{Y}_{nt-1}' \otimes I_{S+1})R,$$

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

$$\widehat{\Psi} \xrightarrow{p} \Psi_0 \text{ as } T \to \infty, \tag{22}$$

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

$$\sqrt{TN}(\widehat{\Psi} - \Psi_0) \xrightarrow{d} N(b(T), \operatorname{Avar}(\widehat{\Psi})) \tag{23}$$

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

$b(T) \to 0$ necessarily if N fixed $T \to \infty$, but possibly not if both N and T are increasing. A consistent estimator for $\text{Avar}(\widehat{\Psi})$ *is*

$$\widehat{\text{Avar}}(\widehat{\Psi}) = \left[(TN)^{-1}\sum_{n=1}^{N}\widehat{Z}_n'\widehat{X}_n\right]^{-1} R'\widehat{\Sigma}R\left[(TN)^{-1}\sum_{n=1}^{N}\widehat{X}_n'\widehat{Z}_n\right]^{-1}, \qquad \textbf{(24)}$$

and $\widehat{\Sigma}$ a consistent estimator for $\Sigma = E[(Y_{nt-2} \otimes V_{nt})(Y_{nt-2} \otimes V_{nt})']$.[24]

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, similar to the suggestion of Anderson and Hsiao (1981). 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 grows. Yet, it will necessarily disappear for fixed N large T.

Next, we wish to recover the structural parameters. Online Appendix Section E.1 establishes that when there is a lone signal $S = 1$, the function $g : \theta \to (\Psi', \text{vech}(\Omega)')$ 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,

$$\widehat{\theta} = g^{-1}\left(\left[\widehat{\Psi}' \ \text{vech}(\widehat{\Omega})'\right]'\right). \qquad \textbf{(25)}$$

Online Appendix Sections E.2–E.4 also detail the identification and estimation of the structural parameters when $S > 1$. The advantage of estimating all parameters together is in added efficiency. But because of the added difficulty of the nuisance parameters, for space considerations, and because signal-by-signal analysis still yields consistent estimates for all parameters of interest, we proceed on a signal-by-signal basis in the remaining analysis.

5. SMALL SAMPLE PROPERTIES

The asymptotic distribution of some elements of $\widehat{\theta}$ depend on the non-Gaussian distribution of $\widehat{\Omega}$. So, conventional asymptotic approximations for confidence intervals are inapplicable. Moreover, analytical corrections for the incidental parameter bias $b(T)$ do not exist. But the fact that $\widehat{\theta}$ is a computationally efficient estimator makes bootstrapping such statistics feasible.

Recall, there is inherently bias in the DPD estimates if N is increasing with T. In order to come to grips with the magnitude of these biases, consider the following values for a hypothetical model with $S = 1$: $\beta = 3$, $\sigma = 1\text{e}-3$, $\rho_y = 0.7$, $\sigma_y = 1\text{e}-3$, $\rho_u = 0.5$, $\sigma_u = 1\text{e}-3$. Using these as data-generating values, the bias in the indirect least squares estimator for θ with variable N and

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

TABLE 2. 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

Notes: Small sample bias in estimator across cross-sectional and time series dimensions.

T is computed by Monte Carlo experiment. The results are summarized in Table 2. 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 about 10% of its true value. So, henceforth, our suggestion is to utilize the bias-corrected estimator $\widehat{\theta}^{\star} = 2\widehat{\theta} - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{(b)}$ with bootstrap draws $\{\widehat{\theta}^{(b)}\}$.[25]

Next, we wish to numerically examine the coverage probabilities of either asymptotic or bootstrap confidence intervals for the bias corrected estimator $\widehat{\theta}^{\star}$. Using again the previous data-generating values, 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 3.[26] In all cases, β is undercovered. 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).[27] Results are described in the bottom pane of Table 3. The results indicate that for a time series dimension 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 are accurate.

Finally, we may also bootstrap percentile confidence intervals for the computed combined measure series $\{\widehat{x}_{nt}\}$ [equation (10)] itself. These are henceforth known as *confidence bands*.[28] 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}

TABLE 3. Monte Carlo: Actual 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							
	$N=1$		$N=3$		$N=6$		$\boldsymbol{N=10}$	
	$T=15$	$T=30$	$T=15$	$T=30$	$T=15$	$\boldsymbol{T=30}$	$\boldsymbol{T=15}$	$\boldsymbol{T=30}$
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: **Bold** indicates coverage probability for all parameters within +/− 5% from actual.

may differ. Is this evidence that officially reported statistics may not be validated? Not just the combined measure, but its entire sampling distribution, is the object of interest: Is the difference between the combined measure and reported output, $\widehat{x}_{nt}-y_{nt}$, *statistically significant*? Level α confidence bands quantify the reasonable range in which true output figures likely lie. We now make use of these statistics in the analysis of the data.

6. RESULTS

6.1. The Asian Financial Crisis: 1997–1999

Recalling the discussion of Section 2, the annualized data set, spanning from 1993 to 2008 ($T=16$) and across all $N=31$ regions, is utilized. On the basis of the results of the Monte Carlo experiments in Table 3, these regions are separated into two groups of 10 and one of 11 with similar elasticities with respect to NO_x emissions before estimation.[29] Group 1 ($N=10$) contains highest elasticities, Group 3 ($N=10$) contains smallest elasticities, and Group 2 ($N=11$) contains mid-range.

Using these groupings, bias-corrected estimates for the structural parameters θ and weightings ϕ are computed. Larger ϕ estimates indicate the given proxy is

TABLE 4. Estimates: Annual sample, 1993–2008

	Group 1 $N = 10$	Group 2 $N = 11$	Group 3 $N = 10$	Pooled $N = 31$
ϕ_L	−0.40	0.01	0.81	0.27
	(−0.81,1.23)	(−1.82,1.91)	(−0.14,2.38)	(−0.38,1.92)
ϕ_N	−0.69	−1.48	0.22	−0.60
	(−1.00,0.81)	(−2.16,−0.06)	(−0.19,1.87)	(−0.94,1.03)
ϕ_F	0.97*	1.24*	−0.79	−0.97
	(0.97,1.02)	(0.78,2.80)	(−2.03,2.26)	(−1.59,0.68)
ϕ_E	0.62*	−0.80	−1.19	−0.76
	(0.07,1.98)	(−1.28,0.64)	(−2.29,0.37)	(−1.23,0.76)
ϕ_C	1.18*	0.95*	1.04*	1.05*
	(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 G.1(a); Group 1 is high NO_X elasticity, Group 3 is low. L: Luminosity. N: NO_2 columns. F: Freight volume. E: Electricity generation. C: Cement production.

relatively more useful in combined measure construction. Signal-by-signal estimates for ϕ with 95% bootstrap confidence intervals are listed in Table 4.[30] Large confidence intervals in all cases disallow us from attributing statistical significance to ϕ for either luminosity or NO_x emissions. Statistical significance may only be attributed to some indices of industrial production, which in the first place are not verifiable.

These results indicate that it is difficult to compute a reliable *point estimate* for a combined measure of economic growth using any signal.[31] This is disappointing in that there is intuitive appeal to having an alternative combined measured of economic growth for use besides GDP. But it is a useful reality check insofar as understanding how much information these signals actually contain, and their robustness in policy deliberations. Signals of economic growth are just that, and they are not correlated with underlying output growth without error. Depending on a point estimate calculated using any one signal leaves the analyst subject to unknown idiosyncratic variation, and should generally be avoided.

However, there is yet usefulness in the computed combined measure. The ultimate purpose of this analysis is to determine whether officially reported output data are supported by other signals of growth. Bootstrapped confidence bands, described in Section 5 and computed using any signal or signals, provide a formal means for determining this. Thus, these signals may be useful regardless of the fact that point estimates are noisy.

As a case in point, in Figure 2, confidence bands across signals are depicted for the four municipalities in the sample.[32] In every case, reported output escapes the confidence bands. So, 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. Why is it that reported

TABLE 5. Estimates: Quarterly sample, 2006 Q1–2013 Q3

	Group 1 $N = 6$	Group 2 $N = 6$	Group 3 $N = 7$	Group 4 $N = 6$	Group 5 $N = 6$	Pooled $N = 31$
ϕ_N	−0.02	−0.27	−1.69*	−1.61	−1.40	−1.58*
	(−2.64,1.20)	(−0.63,1.27)	(−3.91,−3e-3)	(−3.98,1.82)	(−7.03,1.74)	(−3.11,−0.11)
ϕ_E	0.95	−0.30	0.99	1.03	6.95	0.03
	(−1.13,4.09)	(−0.75,1.21)	(−0.02,2.50)	(−1.20,2.98)	(−9.65,11.67)	(−1.45,1.59)
ϕ_C	0.78	−0.51	−0.59	0.41	1.08*	−0.33
	(−0.87,2.40)	(−0.86,1.00)	(−0.99,1.17)	(−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 G.1(b). Group 1 is high NO_2 elasticity, Group 6 is low. N: NO_2 columns. E: Electricity generation. C: Cement production.

output tends to appear too high before the recessionary period, and too low during and after? Recall, officials also face a data smoothing motive [Nakamura et al. (2016)]. If the percentage change in data is overreported for a period of time, it must ultimately be underreported to maintain trend. Thus, we may interpret statistically significant positive-then-negative deviations from confidence bands as evidence of *data smoothing*.

6.2. The Great Recession: 2007–2009

We now consider the later, quarterly time sample which includes the Great Recession and uses NO_2 column data. The first concern is to separate the $N = 31$ regions into smaller groupings with similar structural parameter values. The Monte Carlo experiment summarized in Table 3 indicates that given the longer $T = 31$ length of this data set, a cross-sectional dimension of just $N = 6$ is now likely to lead to correct coverage probabilities. Five groups of this size are defined, beginning with low NO_2 elasticities (Group 1) throughout high elasticity (Group 5).[33]

Quarterly data on luminosity and freight is not available. Table 5, therefore, presents signal-independent estimates of ϕ for the remaining three signals.[34] Once again, confidence intervals tend to be too large to attribute statistical significance to this weighting.

But also again, the more important statistic from the perspective of data validation are the confidence bands arising from each signal. Confidence bands computed using each NO_2, electricity generation, and cement production individually over the quarterly period are depicted for each of the four municipalities in the sample in Figure 3.[35] Most clearly, data from Chongqing are not validated, and judged too high in the aftermath of the crisis, across all signals.

Are the conclusions we draw from this analysis reasonable? Let us consider the case of data from Chongqing. In the midst of the downturn in 2008, the "Chongqing model," brought forth by the city's then Party secretary Bo Xilai, was hailed as an exemplary economic initiative. Confidence bands indicate that the annualized output figures reported just following the NBER dates—peaking at a whopping 65% annualized growth rate in 2010—are not consistent with NO_2

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

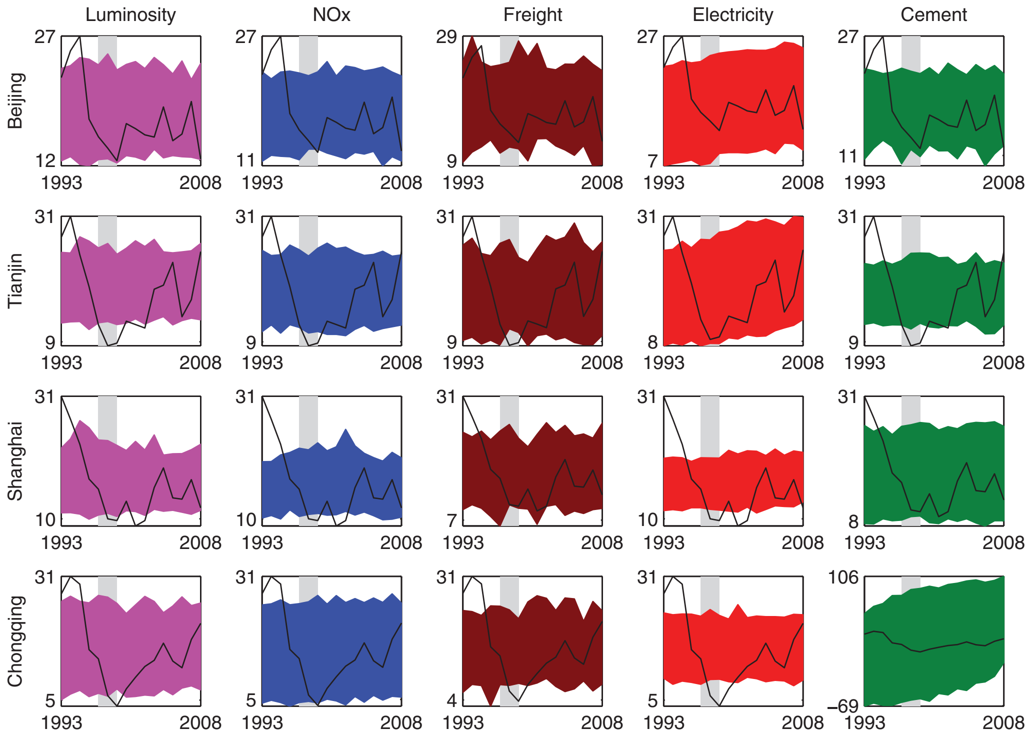

FIGURE 2. Annual 95% confidence bands: Municipalities by signals. *Notes:* Black line: Annual reported % change, regional output. Gray shading: 1997–1999 Asian Financial Crisis period. Colored shading: confidence bands. Confidence bands are computed by bootstrap using each respective Group 1–3 estimates. Beijing, Tianjin, and Shanghai are in Annual Group 1. Chongqing is in Group 2. Point estimates of combined measures are noisy, and do not provide economic intuition, and are thus not depicted. They are available in the data set `annual.mat` using the weightings in Table 4.

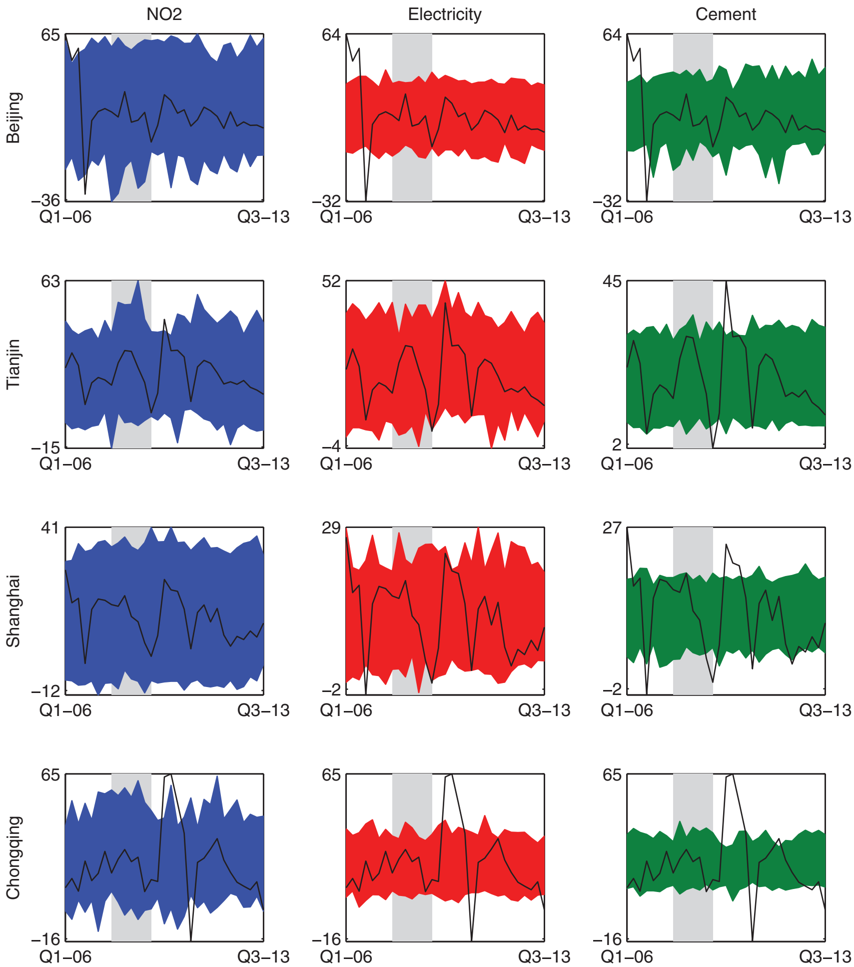

FIGURE 3. Quarterly 95% confidence bands: Municipalities by signals. *Notes:* Black line: Annualized quarterly reported % change, regional output. Gray shading: 2007–2009 Global Financial Crisis period. Colored shading: confidence bands. Confidence bands are computed by bootstrap using each respective Group 1–5 estimates. Shanghai is in Quarterly Group 1. Beijing and Chongqing are in Group 3. Tianjin is in Group 4. Point estimates of combined measures are noisy, and do not provide economic intuition, and are thus not depicted. They are available in the data set `quarterly.mat` using the weightings in Table 5.

columns. In fact, this specific data is also considered suspect by observers outside of this analysis, suggesting this conclusion is robust.[36]

But while data from Chongqing are merely suspect, there are some regions which we now know definitively to have falsified data in this era. These include the northeast provinces of Liaoning, Jilin, and Heilongjiang, which were found

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

to have doctored GDP data during an anti-corruption probe in 2015. Publicly, the official news agency Xinhua cited rampant fabrication of data over past years in a December 11, 2015 report. Online Appendix Figure G.7 shows reported GDP from Jilin (second row, fourth column) was too high during the height of the crisis period. Reported GDP from Liaoning (second row, third column) and Heilongjiang (third row, first column) was marginally too high in the immediate aftermath. Thus, we can also verify ex-post that this methodology works as intended; using only satellite data, we have identified a subset of regions where data is suspect, and may also verify this conclusion is correct for a subset of these regions. At the same time, data *are* validated for 10 out of 31 region throughout the sample, so we are not flatly rejecting all data. Finally, note that in Figure G.7, data from China as a whole are validated (first row, first column). An analyst, using an entirely correct statistical approach, would fail to reject data reliability if blunt national level data were utilized.

7. TWO DIRECT TESTS FOR TECHNICAL CHANGE

How sensitive are these results to the assumptions? One structural assumption made previously was that there is a stable elasticity of the signal with respect to output, β. This was written $S_{nt}^* = Y_{nt}^{*\beta}$ for S_{nt}^* and Y_{nt}^* the unobservable true levels of signal and output. This became $s_{nt}^* = \beta y_{nt}^*$ for s_{nt}^* and y_{nt}^* each percentage change less means in equation (3).

However, if technical change causes substitution toward cleaner technologies over time, then perhaps there is not a stable relationship between the levels of the NO_2 signal S_{nt}^* and output Y_{nt}^*. In particular, perhaps such signals become less responsive to the production of goods and services as technologies evolve and become more environmentally friendly. This real possibility is at the very least worthy of consideration. In this section, we provide two simple and direct tests for this sort of technical change.

First, we consider the implication of technical change for the model's specification. Say the response of the signal to output decays at a regular percentage rate $b \geq 0$. In comparison with the original setup, this would be written $S_{nt}^* = \exp\{-(b/100)t\}Y_{nt}^{*\beta}$, which nests the original setup for $b = 0$. This change generalizes equation (3) to

$$s_{nt}^* = \beta y_{nt}^* - b(t - \bar{t}) \tag{26}$$

for $\bar{t}$ the mean time period. Using (26), the model's new reduced form is solved for in Online Appendix H.

The model's more general solution for b possibly not zero implies there are two direct ways to test for the null hypothesis $H_0 : b = 0$, no technical change. The first way uses the regression,[37]

$$\Delta y_{nt} = b_0 + b_1 \Delta y_{nt-1} + b_2 \Delta s_{nt-1} + \Delta v_{nt}^y, \tag{27}$$

$$\text{Test 1:} \quad H_0 : \ b \equiv b_0/b_2 = 0,$$

https://doi.org/10.1017/S1365100518000056 Published online by Cambridge University Press

TABLE 6. Wald statistic for null: No technical change ($b = 0$)

Period	Sample	Test 1 p-value	Test 2 p-value
Annual, 1993–2008. Signal: NO_x. Contains: *Asian Financial Crisis*.	$T = 16$, $N = 31$.	0.87	0.98
Quarterly, 2006 Q1–2013 Q3. Signal: NO_2. Contains: *The Great Recession*.	$T = 31$, $N = 31$.	0.26	0.29

where Δ is the period-to-period change and Δv^y_{nt} is an MA(2) process. Given this error structure, estimates of the parameters may be obtained using Δy_{nt-3} and Δs_{nt-3}, or any larger lags, as instruments. Given also the structural restrictions underlying these regression coefficients implied by (26) and derived in Appendix H, the null hypothesis may be framed as the simple nonlinear restriction $H_0 : b \equiv b_0/b_2 = 0$. Note, (27) fits within the framework studied by Arellano and Bond (1991), which also uses lagged differences as instruments. Moreover, we need not model any other underlying structural relationships to test this null hypothesis. In particular, we need not directly estimate any of the structural parameters which were previously shown to be subject to small sample biases, and other distortions. Therefore, as in Arellano and Bond's study, there is no reason not to conduct this test using a typical nonlinear Wald statistic and asymptotic results.[38]

The second way of testing for technical change, which also emerges as a consequence of the reduced form derived in Online Appendix H, utilizes the regression,[39]

$$\Delta s_{nt} = b_0 + b_1 \Delta s_{nt-1} + \Delta v^s_{nt}, \tag{28}$$

$$\text{Test 2:} \quad H_0 : \; b \equiv b_0/(b_1 - 1) = 0.$$

Δv^s_{nt} is also MA(2), so Δs_{nt-3} may be used as instruments. Similar to Test 1, the null hypothesis of no technical change may in this case be framed as a set of nonlinear restrictions on the regression coefficients.[40]

Table 6 presents p-values for the null that there is no technical change using either of these tests. In both time samples, across both tests, we may not reject the null hypothesis that there is no technical change. Therefore, the previously held assumption that β is relatively constant across either time sample is formally supported in the data. Given technical change in this parameter is potentially intuitive, how can this result be interpreted? This result in no way precludes technical change in the longer run. Rather, one reasonable interpretation is that in the time samples utilized in this paper, 8 and 16 years, any technical change is small enough for the assumption that β is basically constant to not be a bad one.

Regardless, how would one proceed if they used a longer time sample and found that the null of no technical change, $b = 0$, were rejected? Or if they simply believed these dynamics were important? Equations (27) and (28) not only are the basis of two hypothesis tests, but also make up a generalized version of the main model utilized in this paper (Proposition 1), albeit in differences. Working in differences, the previous analysis could theoretically be entirely repeated, now allowing for $b > 0$. We leave this claim for further study.

8. CONCLUSION

Can officially reported output statistics be externally validated using other verifiable signals of economic growth? This paper has presented evidence that data from China may be either validated, or not, on the basis of international satellite readings of tropospheric NO_2 densities. The results indicate that reported output figures over the Great Recession period are corroborated by satellite readings for many sub-national regions within China. However, reported figures for some areas are not supported, and we now know a subset of these areas to have falsified data due to a 2015 corruption probe, validating the approach. NO_2 column data is freely available at high temporal frequency for areas worldwide. This makes it a useful companion to annual night lights data in the economist's toolkit, particularly with respect to the study of business cycle frequency fluctuations.

A feature of data misreporting in China is that it arises at the sub-national level, so any analysis should be conducted at least in part cross-sectionally. At the same time, elasticities of a given signal with respect to income differ regionally due to differences in production technology, so the time series dimension of the data should also be exploited. This paper has presented a DPD framework which addresses these concerns, and is also generally applicable to areas worldwide. A qualitative finding is that point estimates of combined measures of economic growth are noisy, calling into question the usefulness and/or reliability of point estimates of combined measures of economic growth. Yet, the sampling distribution of the combined measure nonetheless provides a meaningful, formalized statistical basis from which to validate—or fail to validate—reported data.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/S1365100518000056.

NOTES

1. Only recently has there been developed a quarterly macroeconomic data set for China comparable with those commonly utilized in empirical research on Western economies [Chang et al. (2015)].

2. Here and throughout, NBER recession dates for the Great Recession period in the Unted States are used as a proxy to the analogous period in China, since similar dates for China are not publicly reported.

3. Analogs to Figures 1(a) and (b) for each regions depicted in Figures 1(c) and (d) are given in the Online Appendix, Section A, Figures A.1 and A.2. All supplementary tables and figures later labeled "A._" are found there.

4. For example, Alder et al. (2016) have found that point estimates of the effect of Special Economic Zones on growth in China are lower using luminosity-based combined measures, rather than officially reported data alone. In interpreting such results, having a grasp on the range of uncertainty inherent in the combined measure itself is important.

5. A common viewpoint is that data manipulation in China has become more cunning in recent years [Koch-Weser (2013)]. Holz (2014) has argued that previously offered evidences of data manipulation is uncompelling, and that China's National Bureau of Statistics has the freedom to doctor figures in a "virtually undetectable" manner.

6. Liu et al. (2015) estimate that energy consumption in China during the 2010–2012 period was 10% above officially reported statistics.

7. 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)].

8. Daily data are available, but noisy. The monthly average is reliable. We make use of the quarterly average, since we are considering quarterly GDP data.

9. Prominent scientific studies following this approach include Lin and McElroy (2011). Specifically, NO_2 columns are used to constrain NO_X emissions estimates via a chemical transport model relating stocks (columns) to flows (emissions). Purely human-made emissions can be determined independently via a transport model using surface activity data such as land cover, wild fire, and meteorological inputs. 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 [Akimoto et al. (2006)].

10. In fact, NO_2 readings have been used to indirectly infer other more difficult to measure species, such as carbon dioxide, over China [Berenzin et al. (2013)].

11. The evolution of the percentage change in these series for China as a whole is depicted in Figure A.3.

12. Lin and McElroy (2011) previously noted that NO_2 VCDs dissipated considerably during the 2007–9 period. Figure A.4 plots these series over the sample, and once again indicates a clear trough in all series during the crisis period.

13. Figure A.5 depicts estimates by-region for both luminosity and NO_X in the annual sample. These estimates suggest 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 conclusion is found with respect to NO_2 in the quarterly sample, as depicted in Figure A.6.

14. Figure A.9 underscores these regional differences in signal elasticities. It plots the percentage change in electricity generation and NO_2 columns over the NBER dates previously depicted in Figures 1(c) and (d) 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, but any such trend is insignificant.

15. 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. Table A.2 presents the same, excluding luminosity and freight traffic, for NO_2 columns in the quarterly sample.

16. Due to the apparent low correlation of steel production with GDP found in this analysis, and its multiple missing values, for the remainder of the paper, it is omitted in favor of the remaining signals.

17. As a first step of determining whether β is relatively constant across t, one may exploit variability across n to estimate β within each period. In Figure A.7, elasticities for both luminosity and NO_X emissions are estimated for each year in the annual sample. They are statistically indistinguishable from one another over time, for each signal. 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. The fact that NO_X estimates are as constant as luminosity is meaningful insofar as the latter signal is never subject to a time trend; labor and light are non-substitutable. In Figure A.8, the same is shown to hold for NO_2 columns across each quarter in the later sample.

18. Nakamura et al. (2016) use household expenditures to study inflation and consumption at the national level from a time series perspective, but do not consider output directly.

19. 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."

20. Online Appendix Section B provides an example of a simple dynamic stochastic general equilibrium model which corroborates the reduced form assumption in (2).

21. This approach has a long tradition in economic modeling [Sargent (1989)].

22. Consider the situation in which an authority overseeing area n designates an output target. 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 $\varepsilon^\tau_{nt} \sim \mathrm{IWN}(0, \sigma^2_\tau)$ 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 (2). In opposition to this authority, assume that there is an output reporter from area n who manipulates output data y_{nt} exactly to the extent such that it achieves this target in each period: $u^*_{nt} = \tau_{nt} - y^*_{nt}$. Substituting yields $u^*_{nt} = \rho_y u^*_{nt-1} + \varepsilon^\tau_{nt}$. This is the same as the assumed rule of motion for reporting error (6).

23. $\xrightarrow{p}$ denotes convergence in probability and subscript-0 denotes population value.

24. The functional form of the estimator $\widehat{\Sigma}$ is given in Online Appendix D, equation (D.6).

25. See Online Appendix Section F.1.

26. These parameters depend only on Ψ so their confidence interval is easily obtained using the delta method.

27. See Online Appendix Section F.2 for the computation of bootstrap confidence intervals. Following Horowitz (2001)'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 over 40 CPUs.

28. Online Appendix Section F.3 details the computation of bootstrapped confidence bands.

29. 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 emission estimated elasticities $\widehat{\beta}$ listed in Figure A.5 to provide an ordering. The groupings implied by these preliminary estimates are depicted in Figure G.1(a).

30. Estimates for each structural parameter in θ are listed in Online Appendix Tables G.1–G.2.

31. Chen and Nordhaus (2011) also call into question the usefulness of combined measures of economic growth. They conduct a world-wide cross-sectional analysis using luminosity and find that estimates for ϕ are small for any but developing countries, i.e. excluding middle-income countries like China. They do not consider the sampling distribution.

32. Figures G.2–G.6 depict confidence bands across all regions for all signals in the annual sample.

33. The preliminary estimates of NO_2 column signal elasticities $\widehat{\beta}$ for this data set, given in Figure A.6, are used to provide an ordering from which to define these groupings. The results of this objective method of determining groupings is given in Figure G.1(b). 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.

34. Estimates for the structural parameters θ are given in Tables G.3 and G.4.

35. All regions are depicted in Figures G.7–G.9.

36. Xilai was deposed from office and sentenced to life in prison for corruption on September 22, 2013. In the words of his predecessor, Wang Yang in 2009, "Some of our GDP figures sure look rosy." [Liu (2009)].

37. Compare (27) with Online Appendix equation (H.4), first row.

38. See Online Appendix equation (H.5) for the form of Test 1's Wald statistic.

39. Compare (28) with Online Appendix equation (H.4), second row.

40. See Online Appendix equation (H.6) for the form of Test 2's Wald statistic.

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.

Alder, S., L. Shao, and F. Zilibotti (2016) Economic reforms and industrial policy in a panel of Chinese cities. *Journal of Economic Growth* 21(4), 305–349.

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 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. Available on: https://www.wsj.com/articles/SB10001424052748704156304576003222932021228.

Berenzin, E., I. Konovalov, P. Ciais, A. Richter, S. Tao, G. Janssens-Maenhout, M. Beekmann, and E. D. Schulze (2013) Multiannual changes of CO_2 emissions in China: Indirect estimates derived from satellite measurements of tropospheric NO_2 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. NBER Working Paper #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. Federal Reserve Bank of San Francisco Working Paper Series #2015-12. Available on: https://www.frbsf.org/economic-research/files/wp2015-12.pdf.

Ghanem, D. and J. Zhang (2014) "Effortless Perfection:" Do Chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management* 68(2), 203–225.

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*, 56(6), 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 NO_2 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. Available on: https://www.uscc.gov/sites/default/files/Research/TheReliabilityofChina%27sEconomicData.pdf.

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, 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, R. J. Andres, D. Crawford-Brown, J. Lin, H. Zhao, C. Hong, T. A. Boden, K. Feng, G. P. Peters, F. Xi, J. Liu, Y. Li, Y. Zhao, N. Zeng, K. He. (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* 8(3), 113–144.

Neyman, J. and E. L. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16(1), 1–32.

Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49(6), 1417–1426.

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

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, 373–383.

Wallace, J. L. (2014) Juking the stats? Authoritarian information problems in China. *British Journal of Political Science* 46(1), 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.