Is China Fudging its GDP Figures? Evidence from Trading Partner Data

John G. Fernald

(Federal Reserve Bank of San Francisco, United States)

Mark Spiegel

(Federal Reserve Bank of San Francisco, United States)

Eric Hsu

(Federal Reserve Bank of San Francisco, United States)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 2E: Pushing the Boundaries of the SNA

TUESDAY, AUGUST 21, AFTERNOON

Is China Fudging its GDP Figures?
Evidence from Trading Partner Data

John Fernald, Eric Hsu, and Mark M. Spiegel[*]
Federal Reserve Bank of San Francisco

August 2018

Abstract

How can we reliably estimate fluctuations in economic activity for a country with low quality statistics? We propose using imports as a proxy for activity. Imports are one of the best measured components of GDP and are subject to external verification through trading partner statistics. For countries with good statistical systems, imports and measured GDP move closely. But, as expected, the comovement is much weaker for countries with poor statistical systems. We apply this insight to China. By this metric, Chinese statistics, including GDP, have become more reliable over time. But among possible economic indicators that we consider, GDP is merely in the middle of the pack. That said, no single indicator on its own is particularly reliable. Rather, our preferred method for measuring economic activity takes the first principle component of a wide range of indicators. Trying to distinguish among the indicators to find a parsimonious set of indicators within a given sample leads to somewhat worse out-of-sample performance.

Keywords: China, GDP, principal components, structural break, forecasting

J.E.L. Classification numbers: C53, C82, E20, F17

[*] We thank Neil Gerstein, Israel Malkin, Andrew Tai, and Bing Wang for excellent research assistance. Eric Swanson, participants in the CUHK/BOFIT conference on China's Financial Intermediation, and two anonymous referees provided helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of San Francisco or the Federal Reserve System. E-mails john.fernald@insead.edu; ehsu91@gmail.com; mark.spiegel@sf.frb.org.

How can we reliably estimate fluctuations in economic activity for a country with low quality statistics? One approach has been to use light as a check on the statistics (Henderson, et al, 2012; Clark et al, 2017). But light emissions have considerable high-frequency noise, so this approach serves primarily a low-frequency check on statistical quality. It is often of interest to understand cyclical fluctuations as well. China is a clear example where fluctuations are of first-order interest to many observers, including financial markets.

In this paper, we propose using imports as a proxy for activity. Imports are one of the best measured components of GDP and are subject to external verification through trading partner statistics. For countries with good statistical systems, we find that imports and measured GDP move closely. But, as expected, the comovement is much weaker for countries with poor statistical systems.

We apply this insight to China. We find that Chinese statistics, including GDP, have become more reliable over time. But among possible economic indicators that we consider, GDP is merely in the middle of the pack. Nevertheless, no single indicator on its own is particularly reliable. Rather, our preferred method for measuring economic activity takes the first principle component of a wide range of indicators such as electricity or rail shipments. Trying to distinguish among the indicators to find a parsimonious set of indicators within a given sample leads to somewhat worse out-of-sample performance.

Observers of the Chinese economy have long questioned the accuracy of Chinese output figures.[1] Under any circumstances, measuring Chinese GDP would be difficult. China's economy has grown rapidly and undergone extensive structural changes (e.g. Holz, 2008). Many

[1] See Sinclair (2012) for extensive references.

observers further worry that output figures may be distorted, particularly by local and provincial officials in an effort to meet quotas handed down by the government. As a result, many analysts of Chinese economic activity rely instead on alternative, non-GDP indicators.[2]

Skepticism about the accuracy of Chinese data has been shared by prominent Chinese officials. For example, in 2007 current Premier Li Keqiang, was reported as saying that his province's government focused on "alternative indicators," rather than official GDP data (Wikileaks, 2007). Li mentioned three indicators: 1) electricity consumption; 2) the volume of rail cargo, which he suggests is fairly accurately measured because fees are charged for each unit of weight; and 3) the amount of loans disbursed, which may be more accurate because of regulatory oversight. By looking at these three figures, Li said he can measure with relative accuracy the speed of economic growth. Li reportedly said with a smile, "All other figures, especially GDP statistics, are 'for reference only.'"

The challenge in assessing the quality of reported Chinese output figures is to find an independent benchmark to compare with reported data. Following Henderson, et al (2012), Pinovsky and Sala-i-Martin (2016) use satellite data on light emissions to gauge growth in economic activity for a cross-section of countries, including China. China's reported GDP growth rate appears to be exceptionally high relative to its growth in observable light. Clark, Pinovsky, and Sala-i-Martin (2017) focus specifically on China. Although light emissions in the aggregate appear to suffer considerable measurement error, they use the cross-province variation to assess the informational content of indicators available regionally. Nakamura, et al (2014) use household consumption data to estimate Engel curves for China. They find that official

[2] For examples of informal press discussions, see Noble (2015), Sharma (2013), and Bradsher (2012).

aggregate consumption data are too smooth relative to what is implied by household spending patterns.

We argue that inflation-adjusted imports (measured using trading-partner-reported exports) serve as a reliable high-frequency measure of fluctuations. Like measured light emissions, these data are reported externally, so they are not subject to manipulation or mismeasurement by Chinese authorities. However, they should be closely associated with economic activity in China. Specifically, since the data correspond to Chinese imports, they reflect both the use of intermediate inputs for production—an important aspect of China's economy—as well as finished goods imported for final consumption by Chinese residents. As we show below, while the external sector represents only a portion of economic activity, imports co-move very closely with GDP for many economies.

We take this source of information as an indicator of Chinese economic activity and compare movements in externally-reported exports to China to reported GDP, as well as to various combinations of domestically-reported "alternative indicators" of Chinese activity. If we find that movements in externally-reported exports to China are closely associated with movements in reported Chinese data, then we conclude that these series are not spurious, but instead are tracking underlying Chinese activity.

We begin by examining the first principal component of combinations of 14 widely cited and easily available economic indicators, including GDP, produced by Chinese authorities. Our goal is to identify which indicators, singly or in combination, best explain China's externally-reported imports. Principal components estimation proves useful for yielding a parsimonious specification. Some of the individual indicators that we use might be subject to manipulation or systematic mismeasurement; but, if so, our tests would find that they are not closely related to our externally-reported Chinese-import data. Even if the indicators are informative, they might be noisy. By extracting an activity factor as the first principal component, we reduce the idiosyncratic noise in order to focus on the signal.

We begin with a set of 14 potential output indicators, including GDP. We construct the first principal component of all 16,383 possible combinations of these variables and relate them one-by-one to externally reported Chinese imports. This principal-component methodology allows us to focus on a parsimonious relationship and to identify a preferred index of activity. In particular, we relate each combination to externally-reported Chinese imports both in sample and out of sample; the "preferred" index is defined as the best average of in-sample and out-of-sample fit.

** Need to edit below **

Our initial approach compares the information in a small set of potential activity indicators over the full sample of data. The set includes officially reported GDP, the first principal component of all 14 indicator variables, and the first principal component of the three Li indicators. The link between GDP and externally-reported Chinese imports turns out to relatively weak, as either the best-fitting activity factors or the principal component of all 14 indicators correspond much more closely. Moreover, our principal component of all 14 indicators outperforms Li's set. In particular, although electricity and rail freight—two of the Li indicators—are strongly associated with imports, the lending indicator is much less important.[3] Nevertheless, we find relatively little sensitivity to the exact group of included activity indicators in our comparisons of different groups of predictors.

[3] As pointed out by a referee, lending may be a leading indicator of future activity. We do not consider leads and lags because we would end up with too many potential combinations. However, the lending variable in particular does better when our data are restricted to the post-2008 period. (??).

The initial results do suggest that the accuracy of reported GDP—as well as the activity factors—has improved over time. Improvements could reflect rising dissatisfaction by Chinese officials about the quality of their statistics. For example, Chinese officials have increasingly and openly discussed their concerns, as the Li quotation suggests. This dissatisfaction could only have increased with the onset of the global financial crisis, as successful implementation of the aggressive counter-cyclical measures adopted by the Chinese central government during the crisis required accurate assessments of prevailing economic conditions.

Nevertheless, when we estimate factor models, we find that it is preferable to use a long sample to estimate the factor structure. We reach this conclusion by doing out-of-sample tests of predictive power. In particular, for each of the 16,383 possible combinations of individual indicators, we look at whether the out-of-sample fit is better if the factor structure was estimated over a long sample (starting in 2000) or a short sample (starting in 2008). In 94 percent (CHECK) of cases, the out-of-sample fit is better when the factor structure was estimated on samples that began in 2000. Intuitively, there is a tradeoff between bias (if the factor structure has changed) and precision (if the sample is too short). This finding that factor estimation should be done with a long sample is consistent with the recommendation of Stock and Watson (20xx).

[Of course, each of our alternative indicators by construction focuses on specific areas of the China economy. As such, it is plausible that the time series of Chinese imports does not follow those of many of our alternative indicators exactly. But GDP is supposed to be the broadest measure of economic activity. By including GDP as one of the indicators, its variation is included in our measures. However, adding GDP or not adding GDP makes very little differences to the explanatory power of our preferred principal component indices.

Our emerging picture seems to be one where reported GDP is better at predicting Chinese activity as proxied by externally-reported import data than it used to be. GDP adds modestly to the accuracy of the fit of our best combinations of alternative indicators, and these indicators also do a good job of predicting GDP. Still, these improvements are modest at best, and there are combinations of domestically-reported alternative indicators of Chinese economic activity that predict Chinese imports more accurately when GDP is excluded.]

The remainder of this paper is divided into six sections: Section 1 discusses the relationship between imports and measured GDP, and how this relationship depends on statistical capacity. Section 2 describes our data and methodology. Section 3 argues for using the full sample to estimate the factor loadings, despite evidence that the quality of statistics has improved over time. Section 4 shows our main results. [Section 5 conducts a number of robustness tests, including the examination of relative out-of-sample performances.] Section 6 concludes.

1. Imports as a measure of activity

The challenge in assessing the reliability of different economic indicators is that we need a benchmark that is highly correlated with true activity but is not, itself, subject to manipulation. In this section, we document that a country's imports fit that bill: Import growth moves closely with GDP growth for countries with relatively reliable statistical systems.

Why would we expect imports to be one of the best measured components of the national accounts? First, the number of importers (and import locations) is typically modest, which makes measurement more manageable. Second, countries have an incentive to measure imports accurately for tariff purposes. Third, imports are subject to external validation through trading partner data. So deviations from accuracy are verifiable.

In countries with poor statistical systems, we would expect the relationship between imports and measured GDP to deteriorate simply because measured GDP becomes less accurate. The reduced accuracy of measured GDP should then reduce its correlation with imports. In contrast, for the reasons noted above (including the external verification), there is little reason to think that the correlation between imports and *true* economic activity deteriorates.

To assess these conjectures, we look at cross-country data to see the relationship with statistical capacity. We use data on 165 countries from the Penn World Tables (release 9.0). For each country, we calculate the correlation of growth in real imports and real GDP from 1990 to 2014, using national source data (the data that underlie the more-frequently used purchasing-power-parity data). For each country, an earlier version of the PWT (release 6.1) had a judgmental measure of statistical capacity, which ranked the countries from A (highest) to D (lowest).[4]

Table 1 relates the import-GDP correlation across countries to statistical capacity and other control variables. The control variables are initial income per capita (GDP per capita in 1990, in international dollars) and openness (nominal exports plus imports relative to nominal GDP, in local currency). Income per capita could, independently, be associated with| the correlation between imports and GDP. For example, the structure of the economy—say, goods relative to services—might systematically be related to the level of income. Since many low-

[4] See Dawson et al., 2001, for a discussion of how this measure is constructed. They note that for some issues in the growth literature, accounting for statistical capacity is important. Henderson et al (2012) use a different measure of statistical capacity from the World Bank, but that one is *only* available for developing economies, not for the full universe of countries.

income economies have low statistical capacity, we want to be sure that statistical capacity is not simply proxying for income. We might also expect that very open economies might have a weaker relationship between imports and GDP, because a lot of trade is simply passing through the port, and may be responsive to *foreign* rather than domestic conditions. (We use 1990 values for openness and GDP per capita, but using average values is very similar).

The first column shows that countries with poor statistical capacity (C or D) do indeed have a notably lower correlation between growth in GDP and in imports. The constant term shows that a country with an A-rated statistical capacity (the omitted category in the regression) has a correlation that is close to 0.8. For these countries, imports move quite closely with GDP. However, for a country with D-rated statistical capacity, the correlation falls to 0.35 (the 0.788 constant term minus the 0.426 coefficient on the "D" statistical capacity dummy).

The second column omits statistical capacity but adds GDP per capita and openness. The unconditional correlation is, in fact, positively and significantly related to GDP per capita. That is, richer countries have a higher estimated correlation. The correlation is insignificantly negatively related to openness. Note that the explanatory power is much lower with these controls than with the statistical capacity measures (an R^2 of 0.08 rather than 0.22 in column 1).

The third column shows that when all of the controls are included together, it is statistical capacity that is most strongly associated with the correlation. The coefficients on statistical capacity as well as the R^2 are very similar to those shown in the first column. GDP per capita and openness are insignificant. These results suggest that, in column 2, GDP per capita was proxying for statistical capacity rather than the reverse.

Thus, the statistical evidence suggests that, in countries with relatively good statistical systems, GDP and imports move closely with each other.

For comparison, the United States has a statistical rating of A and a correlation of above 0.9. China has a statistical rating of D and a correlation of about 0.6. China's correlation is above its expected value, conditional on its statistical capacity rating of D. However, it is below what would be expected for a country with an A or B rating. The next section considers what we can learn about the quality of China's various statistical releases.

We conclude this section by noting the caveat that even imports are an imperfect and perhaps noisy measure of economic activity in China. Moreover, structural change might mean that the relationship between this indicator and other indicators has changed over time. For example, in annual U.S. data for the 30 years (1986-2016), import growth is more highly correlated with growth in goods (0.9) than in structures (0.6) or services (0.5). That said, the correlation is highly significant even with these different components of activity. And, even if it noisy and imperfect, there is little reason to think it is biased.

In our empirical work for China, we consider whether the relationship has changed by looking at time variation. A priori, structural change seems like it could cause relationships to attenuate over time, given that China's service sector has grown in importance relative to the traded sector. But, empirically, we find that the relationship has improved over time.

2. Chinese data

Our goal is to use the insight from the previous section about the information content of imports to develop a reliable indicator of activity in China. This section discusses what Chinese data we use to achieve that goal.

A. Measuring China's imports

For any country where the accounts are suspect, including China, there is a question of whether the import statistics themselves are accurate. As noted already, a key advantage of imports is that they can also be measured using trading-partner exports. For both economic and statistical reasons, we combine exports to China and Hong Kong for these purposes. Economically, many of the goods that are exported to Hong Kong from non-China sources are destined for the Chinese mainland.[5] Statistically, authorities in, say, the United States may plausibly have changed the degree to which they are able to track the ultimate destination over time—that is, a good that previously would have been recorded as an export to Hong Kong might now be recorded as an export to China. Using the combination of Hong Kong and China makes the data more comparable over time.[6] (In the robustness section, we plan to check whether it makes any difference which source we use.)

For our main analysis, we continue to use trading-partner-reported data, since it is immediately clear that measurement error in this indicator should be independent of the measurement error in China-source economic indicators. This source of data on China's imports is not controlled in any manner by Chinese authorities. (Henceforth, when we refer to imports, it's always as reported by trading partners.) Trading-partner governments have no apparent incentive to misrepresent their trade volumes with China. Of course, the rapid growth of trade with China could still cause some measurement challenges for these countries. However, these data still have the advantage of being measured at foreign ports. Moreover, while Chinese trade is growing as a share of total trade for these countries, overall trade is not growing nearly so fast.

---

[5] For example, in 2016 over US$400 billion in goods were re-exported through Hong Kong from and to the Mainland (https://www.tid.gov.hk/english/aboutus/publications/factsheet/china.html).

[6] Fernald, Edison, and Loungani (1999) argue for combining Hong Kong with China. We confirm in the data appendix that imports by China and Hong Kong imports (henceforth referred to as "China's imports") move very closely with its trading-partner-reported exports (see Appendix Figure A2). The trading-partner data line up better for China plus Hong Kong than for China alone.

So tracking trade volumes, including those destined for or originating from China, is less challenging.

Using the IMF's Direction of Trade Statistics (DOTS), it is straightforward to measure trading-partner exports to China and Hong Kong. We obtain similar results to those reported later in this paper when we use narrower sets of countries—such as exports from the United States, the Euro Area, and Japan. Because imports represent intermediate inputs and final consumption or investment goods, they are likely to be correlated with overall activity. To calculate real imports, we deflate with a China-specific export deflator, described in the data appendix.[7]

B. *Individual data series*

From Chinese-source data, we identified 14 potential activity indicators on the basis of data availability and a priori plausibility—GDP plus 13 non-GDP variables. The 14 indicators are all available from the beginning of our sample (the beginning of 2000), and were downloaded from CEIC Asia. Examples include electricity use, industrial production, rail freight, and new property construction. The indicators are described in the data appendix and also listed in the tables in the next section. Although GDP is of independent interest, for our main purpose ("what is the best index of activity in China") we consider GDP as just one of a list of possible indicators to examine.

[To determine the number of indicators used, we added indicators one at a time and examined whether the additional indicator improved our goodness of fit, as measured below. We

[7] We use U.S. product-level export prices, weighted by the product-level imports of China and Hong Kong. U.S. measures of prices are considered relatively reliable; and any biases are likely to be unrelated to economic conditions in China. We assume that the product-level weights from China and Hong Kong sources are relatively accurate. In any case, the bias for the overall deflator is likely to be small, even if the weights are unreliable.

found that the 15th indicator provided little improvement in goodness of fit between the best-performing indicator combinations and our Chinese import measure.[8]]

To control for seasonal factors, we use all variables in four-quarter changes.[9] In principle, we could use the Census X-12 program to control for seasonality. For our purposes, we prefer the simple and transparent year-over-year change. In addition, Wright (2017) raises questions about the reliability of the X-12 procedure.

Before doing any statistical analysis, we follow Stock and Watson (2016) and detrend all individual indicators with a biweight filter. The biweight filter is essentially a smooth two-sided filter that becomes increasingly one-sided at the end points. The reason for filtering is that individual series have different trends—which can be misleading, since our principal component indices will attempt to fit those trends as well as the fluctuations.

For example, without filtering, Chinese GDP growth in 2016 was about 2 standard deviations below its longer sample mean. That growth rate was comparable to its level in the downturn of the Great Recession. Yet, while growth was relatively slow, that appeared to be a trend development, rather than as an indication of slow cyclical growth. The data appendix shows the raw individual indicators, their estimated trends, and the detrended indicators. In all cases, the detrended data have been normalized to have mean zero and unit standard deviation.

A question is what bandwidth to use for the filter. For U.S. data, Fernald, Hall, Stock, and Watson (2017) use a biweight filter with bandwidth of 60 quarters before estimating a factor model. That filter yields relatively smooth trends, which are not too sensitive to end-point issues.

[8] The 15th indicator considered was the inward flow of foreign direct investment.

[9] Many of the series are available monthly, but we convert all data to quarterly terms. Doing so facilitates comparisons with quarterly GDP data, smooths some high-frequency measurement error, and avoids problems with the timing of the Chinese New Year (which sometimes occurs in January, sometimes in February, and sometimes overlaps both).

Even there, however, it is clear that the filter is too smooth to capture the trend for some U.S. series, such as productivity.

For China, the changes in trend growth since 2000 are much sharper than for the United States, and a more responsive filter appears to fit the data better. For this reason, we use a filtering parameter of 24 quarters, which is flexible enough to fit the trends reasonably well. Despite this flexibility, the trend is not overly sensitive to end points, so revisions to real-time estimates of the trend are not too large. For example, we estimated the trend through 2009—when the cyclical deviation from any trend was clearly very large—and compared the real-time trend to the revised trend. While there were revisions, they were not extreme. It is important to note also that the main results do not appear driven by the choice of filtering.

C. *Principal components*

To identify the "best" index of China's cyclical fluctuations, we focus on principal component indices from alternative sets of potential indicators. The reason is that individual indicators that move closely with China's imports in any given sample sometimes fit much less well out of sample. Using principal components of a set of indicators helps minimize this problem.

For example, an extremely misleading approach to using the individual indicators would arise if one simply regressed China's imports on all 14 of the indicators plus GDP. Such a regression has a high R^2 even though, because of multicollinearity, few of these indicators are statistically significant. However, because of overfitting, this approach performs very poorly out of sample relative to a more parsimonious specification.[10]

[10] For example, we regressed the import data on all 14 indicators from the start of our sample until end-2012 and predicted out-of sample thereafter. For comparison, we also regressed the data on the first principal component of these indicators, as well as the first principal component of the three Li indicators. As expected, the regression with

Principal components help minimize the risk of spurious fit by capturing the key common information in the indicators — known as "activity factors" — in a parsimonious way. Principal components are defined by the property that all factors (or components) are orthogonal, with the first component explaining the maximum variation in the included data, the second one explaining the second most variation, and so forth.

One extremely parsimonious set of indicators is the index that includes GDP alone—a single indicator. The a priori justification is that GDP is, in principle, the broadest measure of economic activity. At the other extreme, an a priori reasonable benchmark index is the first principal component of *all* of our individual indicators, including GDP. That benchmark is agnostic about which indicators are informative or uninformative, and whether that informational content has changed over time.

Figure 1 shows selected indicators along with real exports to China. All variables represent year-over-year growth rates and are normalized to have mean zero and unit standard deviation. The indicators shown are electricity, which is often taken as a proxy measure for activity in China; the first principal component of all 14 indicators ("all indicators"); and GDP.

Clearly, the all-indicators activity factor and imports are very highly correlated. For example, during the global financial crisis, both series drop about 3 standard deviations below their respective means. In the recovery, both series rise to above 2 standard deviations above their means. Thus, reassuringly, imports and the activity factor tell the same story about economic activity.

all 14 indicators individually had the lowest (best) RMSE in sample, 0.34 versus 0.52 and 0.65 for the first principal component of all 10 and the Li indicators respectively. However, the regression with all 14 indicators included had A higher RMSE out of sample than the first principal component, 0.57 versus 0.49. The out of sample results for the Li indicators were modestly worse, with an RMSE of 0.58.

However, the relationship of reported GDP with either the activity factor or imports is less strong. The correlation is still positive and significant, but GDP rises more prior to the crisis than either imports or the activity factor, and falls less during the crisis.

A key goal of the sections that follow is to identify which indicators (including GDP) are particularly informative, in terms of correlation with our externally-reported import indicator. It is possible that a more parsimonious index will be an even better index of economic activity. The reason is that adding additional variables could add variation that is largely idiosyncratic and unrelated to imports and overall economic activity. As a result, the first principal component of a larger set of indicators might be less accurate as a measure of activity because it is trying to explain that idiosyncratic variation as well as the systematic variation that matters for imports.

Concretely, we proceed by constructing the first principal component of all possible subsets, other than the null set, of these 14 variables (GDP plus the 13 non-GDP indicators), considering a total of 16,383 combinations. For example, 14 of the combinations have just a single indicator (each of the 14 variables); at the other extreme, one combination uses all 14 variables at the same time (our "all indicators" factor plotted in Figure 1).

For each subset, we then regress growth in Chinese imports from China's top ten trading partners on the first principal component as well as exchange rate values (which plausibly affect import levels independently of output). Our baseline specification is thus

$$\Delta^4 m_t = c + \beta_i PC_{1t} + \gamma \Delta^4 RMB_t + \eta_t \qquad (1)$$

$\Delta^4 m_t$ is reported quarterly growth in real Chinese imports from (measured as real exports to China by) the United States, the euro area, and Japan; PC_{1t} is the contemporaneous value of the first principal component from the year-over-year growth in the chosen set of alternative indicators of Chinese economic activity; $\Delta^4 RMB_t$ is the four-quarter change in the renminbidollar exchange rate; and η_t is an error term. We estimate with ordinary least squares and show Newey-West standard errors that allow for heteroskedasticity and autocorrelation.

The reason for controlling for the exchange rate is to control for non-activity factors that might affect imports. Conceptually, demand for imports in China could depend on two factors. The first is direct final demand for consumption, investment, or government. The second is imported intermediates that are destined for re-export after some further processing. Some but not all of the activity included in the second category reflects activity being done in China. But some reflects derived demand from activity taking place in countries to which China exports.

Now suppose that the RMB depreciates. That might directly lead to a shift of domestic demand towards domestically produced goods and away from the now-more-expensive foreign goods. It also makes imported intermediates more costly, which might affect the incentives to use those parts rather than domestically produced ones. By controlling for the exchange rate, we allow the regression to control for these non-activity channels.

Of course, the concern with a parsimonious index is that in-sample (IS) and out-of-sample (OS) fit might be quite different. This is most obvious for the combinations that comprise only a single indicator. Purely by chance, some indicator might happen to move closely with China's imports during a given sample. To guard against this concern, we seek combinations of indicators that are not only a priori plausible but that perform well both in-sample and out-of-sample.

3. Initial Analysis: What sample to use, given possible structural change?

China's economy has changed dramatically in recent decades. That might point towards wanting to focus on a relatively recent time period. However, the literature on forecasting in the presence of structural change (e.g., Pesaran and Pick, 20xx) finds, in Monte Carlo simulations, that the loss of estimated precision from using a shorter sample period can be more important than the bias in the true coefficient caused by structural change. In this section, we find that it is preferable to use a relatively long sample period to estimate the econometric relationships.

Conceptually, the relationship in equation (1) could change in two ways that would affect the reliability of the relationship. First, the coefficient β_i on some particular principal component index (PC_i) could change over time, perhaps reflecting structural change. Second, even if the β_i do not change, the statistical fit of the relationship could change if the variance of η_t changes. (η_t could, of course, reflect idiosyncratic noise in imports or in the principal components.) In the second case, an indicator might be unreliable even if it is an unbiased estimate of activity.

First, Bai-Perron (2003) break tests find no evidence of structural changes in the parameters β_i . Specifically, in results not shown, we estimated equation (1) with three different indices (the PC_i's): GDP growth, the all-indicators index, and the Li index (the principal component of electricity, bank lending, and rail cargo). With import growth as the left-hand-side variable, as in equation (1), there is no evidence of instability in the regression coefficients. Similarly, if we use GDP growth on the left-hand side, there is no evidence of instability in the relationship with the all-indicators or Li indices.

Second, to look at the overall evidence of change in the statistical relationship, we look at the R^2 from 16-quarter rolling regressions. The R^2 shows the ratio of "explained" variance $\hat{\beta}_i PC_{1t} + \hat{\gamma}\Delta^4 RMB_t$ to total variance; correspondingly, it rises if the idiosyncratic variance η falls relative to the explained variance of China's imports. In contrast, the break tests look more narrowly for subsample changes in β. A good indicator not only has a stable relationship with activity, as the break tests indicate, but also explain a lot of the variation in imports (high R^2).

Figure 2 shows the rolling R^2 estimates from estimating equation (1) with selected indices of activity. In all cases, the left-hand-side variable is growth in China's imports. The right-hand-side variable is the indicator shown (GDP, the first principal component of all 14 indicators, and the first principal component of the three Li variables). The regressions also include the real exchange rate, though results are little affected by omitting it.

The important takeaway from the figure is that all of the indices, especially GDP, fit very poorly prior to 2008, but fit much better thereafter. Some of this is just the sharp decline and rebound during the Great Recession itself that was common to imports and all of the activity indices. But it is not just that. Even in samples that end 2014 or after (so they are not affected by the sharp downturn of the Great Recession), the indices fit better than before 2008. Other than GDP, the indices fit much better than before 2008.

Note that the all-indicators indicator that includes GDP has virtually identical performance as the index that includes all indicators *other* than GDP. Excluding GDP performs marginally better at a few points early in the sample. At the end, even as GDP on its own deteriorates somewhat, including GDP with the other indicators looks essentially the same.

In light of the break tests and rolling regression R^2's, it is unclear what sample to use. The break tests do not suggest an econometric reason for shortening the sample, but the R^2's suggest that the relationship is much closer after 2008.

To address this uncertainty, we perform out-of-sample exercises after estimating the relationship in equation (1) for different in-sample periods. We find that using the full sample, starting in 2000, unambiguously does better for explaining exports to China out of sample than starting in 2008. Specifically, for each of the 16,383 combinations of indicators that we consider, we estimated equation (1) for the 2000-2013 period and then the 2008-2013 period. Then we looked at how well the estimated relationships fit in the 12-quarters of 2014 to 2016. In only 6.9 percent of the combinations did the out-of-sample (2014-2016) estimates have a lower RMSE when we used the 2008-2013 period for estimation rather than the 2000-13 period. (When we ended the in-sample estimation at the of 2014, only 8.7 percent of the tuples did better with the 2008-2014 period for estimation, so results are little affected.) In other words, in the vast majority of cases, we do better by using the full sample for estimating the relationship between exports to China and indicators.

Table 2 shows a subset of these results. Specifically, it shows the in-sample and out-of-sample fit for the 14 individual indicators we use plus several combinations. The in-sample root mean squared errors (RMSEs) are not comparable, because they correspond to different sample periods. But the out-of-sample RMSEs all correspond to the 2015-16 period; they differ only because the coefficients were estimated over different periods. For all of the individual indicators other than exports, the RMSE is lower when the relationship is estimated over the longer 2000-2014 period. This is also true for the Li index and the all-indicators index.

Of course, the improved fit using the longer estimation sample is also apparent in average the out-of-sample RMSEs for 2014 to 2016, as shown in the bottom line of Table 2. When the in-sample period was 2000-2013, the median out-of-sample RMSE across all tuples was 0.48. But when the in-sample period was the shorter 2008-2013 period, the median out-of-sample RMSE rises to 0.84 (the figures are 0.45 and 0.57 if the in-sample period ends in 2014). Concretely, for the all-indicators tuple, the out-of-sample RMSE was 0.35 when the relationship was estimated over the long sample; but, when the relationship is estimated over the shorter 2008-2013 sample, the out-of-sample RMSE rose to 0.50.

Before moving on, we note that in Table 2, the first nine individual indicators (through the consumer sentiment index) all have in-sample RMSEs that are less than one. Because the import index on the left-hand-side is normalized to have a unit standard deviation, the regression no explanatory variables would have an RMSE of one. Thus, the first nine variables reduce the RMSE relative to omitting the variable, whereas the bottom five actually do worse. The divide is particularly sharp with the 2008-2014 in-sample period. We also note that while there is some tendency for indicators that do well in-sample to also do well out-of-sample, with low RMSEs, the ranking is far from one-to-one. For example, property performs relatively poorly (8^{th} best) in the 2000-2014 estimation period, but is the single best individual indicator out of sample.

In sum, the break tests and the in-sample/out-of-sample tests both suggest using a long sample. We follow that approach in the next section. The rolling estimates do find that even GDP is more informative after 2008. That said, these do not suggest that one should focus solely on GDP. Rather, they suggest that it is still preferable to use a wide range of indicators. In the next section, we consider whether we should use a parsimonious set of indicators.

4. Results for individual indicators

Given the superior performance of estimates fitted over our full sample, we next investigate the relative performances of a variety of alternative activity indicators over the full sample. Tables 3a and 3b summarize our estimation results. The estimated parameter values are not interesting per se, so we do not show them. We instead focus on (i) indicator names and sets; and (ii) fit as measured by RMSE.

We are interested in both in-sample performance and out-of-sample prediction. As we demonstrate below, the relative quality of fit of combinations of activity indicators in and out-of-sample can differ markedly, and it seems plausible that the structural changes that have recently taken place in the Chinese economy may lead some indicators to erroneously indicate strong or weak performances out of sample.

It is unclear in the literature how to weight the relative importance of in- and out-of-sample fit. In response, we characterize performance in terms of the average RMSE in and out of sample. There is also no consensus about the share of a time series sample that should be allotted to the estimation of weights on activity indicators in sample and that allotted to gauging the performances of these indicator combinations out of sample [e.g. Hansen and Timmermann (2012)]. In response, we follow the literature [e.g. Kelly and Pruitt (2013)] in considering the relative performances of our activity indicators using alternative sample splits. We report results in the text with two-year out-of-sample horizons, and then compare these to result with a three year out-of-sample forecast horizon in the appendix.

We first consider the relative performances of indicator combinations by pursuing a sequential approach, as follows. We start by identifying the best-performing single indicator in terms of average RMSE. Next, we then find the best-performing pair of indicators that includes that best single indicator plus one additional indicator. We then find the best-performing combination of three indicators that includes the first two, and so on.

Table 3 shows these results. It can be seen that our best-performing individual indicator is electricity, with an average RMSE in and out of sample of 0.61.

We get a substantive improvement in average performance by adding a second activity indicator, "Property," as the average RMSE drops from 0.61 to 0.53. This is driven by a substantive improvement in out-of-sample fit, as the out-of-sample RMSE drops from 0.56 to 0.36, while the in-sample RMSE actually increases from 0.66 to 0.71.

We obtain further improvement in average RMSE by going to three indicators with the addition of industrial production. This drops the average RMSE to 0.529, but in this case the improvement is driven by a drop in in-sample RMSE from 0.71 to 0.65, while the out-of-sample RMSE increases from 0.36 to 0.41.

None of the remaining individual indicators improved fit as a fourth indicator, as our average RMSE for the best combination, which was achieved with the addition of the floor space indicator had a modestly larger average RMSE than our best-performing three-indicator combination (0.544 vs. 0.529).

Overall, it can be seen that average RMSEs are not monotonically declining with the addition of more activity indicators under the sequential method. We achieve the best average performance with a combination of 9 activity indicators, including electricity, property, industrial production, floor space, consumer expectations, exports, rail, GDP, and government revenue. This combination of 9 indicators achieves an average RMSE of 0.46, with a 0.63 RMSE in sample and a 0.29 RMSE out of sample.

There are a number of notable patterns to the sequentially chosen sets of indicators. First, our best set of 9 indicators is not much better than a number of the other combinations in terms of RMSE, including the all indicators set, which has an average RMSE of 0.50. The conclusion is that one cannot say a priori that one could do better with a more parsimonious set of indicators than one would do with the all indicators set. Still, there appears to be some room for improvement even using this relatively simple algorithm for choosing indicator combinations.

Second, the indicators differ widely in their relative in and out of sample performances. Our best-performing combination of indicators in-sample is the 6-variable combination, which excludes the rail, GDP, and government revenue indicators from the best overall 9-indicator combination. Our best-performing out-of-sample combination is indeed the aforementioned best overall 9-indicator combination.

Finally, note that GDP is included in our best-performing 9-indicator combination, but it is only the seventh indicator to be introduced. This is consistent with our finding above that among the individual indicators, GDP ranks only modestly above average, both in and out of sample. As such, we find that GDP is informative, but can do a much better job of predicting Chinese economic activity when combined with other informative indicators.

To examine the robustness of our results to alternative sample splits, we use the sequential method again in the appendix to evaluate best-performing indicators over a three-year out-of-sample horizon. Our results are qualitatively similar to those we obtain in Table 2. The best-performing set of indicators is identical to the set of 9 activity indicators we obtained using a two-year out-of-sample horizon. However, the path taken to reach this sample is different, as GDP is now introduced as an activity indicator into the 4-indicator variable set.

The sequential method is computationally simple, but does not necessarily yield the best combination of activity indicators. To obtain those combinations, we examine all possible combinations of the individual activity indicators for each number of potential indicators from one to fourteen. We then again choose the set of indicators for each number with minimum average in-sample and out-of-sample RMSE.

Our results are shown in Table 4. We continue to observe wide discrepancies in relative in and out-of-sample performances for the various activity indicator combinations. Tautologically, electricity remains our best-performing individual indicator. However, our best combination of two indicators does not include electricity at all, and instead includes government revenue and rail. Relative to electricity alone, this combination exhibits 0.50 RMSE in sample, but achieves a very low 0.22 RMSE out of sample.

Our best combination with three indicators adds industrial production, which improves both in and out-of-sample performances, and lowers average RMSE to 0.44. Our four-indicator combination adds electricity and exports, but removes industrial production. This indicator combination of the electricity, exports, government revenue, and rail indicators adds a bit to out-of-sample RMSE, but improves in-sample RMSE sufficiently (it is lowered to 0.58) to lower average RMSE to 0.42. Before rounding, this four-variable indicator's RMSE is 0.423, which just beats our best-performing five-variable set of indicators, which includes all of the indicators in our best-performing four-indicator set, and adds retail sales. This five-variable indicator's performance is essentially identical, with an average RMSE of 0.424, as is the best-performing six-variable indicator combination which includes all of the indicators in our best-performing five indicator combination, plus air passengers, and achieves an effectively-identical average RMSE of 0.427.

Second, our overall best set of indicators is not only relatively parsimonious, but also fails to include GDP. GDP is not chosen as a component of the activity indicator set until we are observing 9 indicators. Moreover, this set with GDP included (which is the best-performing set with GDP included as well as the best-performing set we chose using our sequential method) has an average RMSE of 0.44. This is neither notably worse than our best-performing set of four indicators (0.42 RMSE), but also not much better than one could do by simply including all of the indicators, which results in a 0.50 RMSE.

As in our sequential approach, our best indicator approach results are qualitatively robust to an extension to a three-year horizon. There are a wide variety of indicator combinations that essentially yield the same average RMSE performances, ranging from combinations 5 to 10 indicators. Our “best” performing index includes 7 indicators: electricity, property, industrial production, rail, retail sales, consumer expectations and government revenue. Moreover, as in our two-year horizons results, GDP is introduced as an indicator relatively late, in this case only when using 12 or more indicators. Again, GDP has information, but does best in combination with a wide variety of other indicators.

Finally, it is reassuring that the 9-variable indicator combination that we chose over a 2-year horizon also does relatively well for a three-year horizon. While this is not the best-performing indicator over a three-year horizon, it does well relative to the chosen 9 variable combination for this horizon, which only differs in composition from the best-performing 9 variable combination for a three-year forecast horizon by its inclusion of exports and GDP, and its exclusion of consumer expectations and fixed asset investment. Moreover, its performance is superior to that of the all indicator combination, with an average RMSE of 0.49 vs. 0.54.

Overall, the comparison of Tables 3 and 4, along with the longer-horizon sets of indicators in the appendix highlight a number of results: First, while we observe some discrepancies, the qualitative set of individual indicators that perform best is relatively stable. Indicators such as electricity, rail, and industrial production seem to be ones that one would always want to include. Second, the all-indicators combination does well enough that one could rationally chose to use that indicator, letting the data speak solely through the weights chosen in generating the principal component for the activity index.

Indeed, while we were able to construct more parsimonious combinations that outperformed the all indicators activity index, this discrepancy should not be exaggerated. Figure 3 displays the predicted values of the all indicator index, as well as the best overall index, the best index generated by the sequential process, and the best in-sample indicator, against reported Chinese GDP and exports to China and Hong Kong.

It can be seen that all of the activity indices tend to move closely together, and (by construction) tend to move closely with movements in Chinese imports. Still, during episodes where there is a divergence between reported Chinese GDP and other-country reported exports to China and Hong Kong, as in 2007 and 2011 (CHECK DATES), we do see movements in the activity indicators away from Chinese imports and towards reported Chinese GDP.

This can be seen in more detail through Figure 4, which displays the all indicators combination and the all-indicators-except-GDP combination against reported GDP and Chinese and Hong Kong imports. These series tend to closely track each other, except during periods where the GDP series diverges notably from the import series, as in the pre-crisis years. As the fit between Chinese imports and Chinese activity is not perfect, we prefer to include reported GDP as an indicator, which leads to more intermediate predictions during these divergent episodes.

6. Conclusion

[To be added]

We conclude with several caveats. First, imports are an imperfect measure of activity and may underweight certain activities, notably services and other non-tradable sectors. Still, imports are very highly correlated with our preferred activity factor. And that factor includes both relatively narrow indicators (like rail freight and, possibly, raw materials) and broader ones (such as air passenger volume and retail sales). Moreover, even if imports or the activity factors are imperfect, there is no reason to think they are necessarily inferior to GDP alone.

Second, even for the pre-2008 period—when GDP is a poor fit of our Chinese economic activity proxy—we cannot say for sure whether GDP was manipulated, or merely limited in its coverage. If manipulation was rampant, we would expect it to be more prevalent during periods of exceptionally high or low economic activity, as data might be changed to more closely meet trend output goals. There appears to be some evidence of that here during the global financial crisis, but we cannot say whether the level and variability in GDP are accurate.

Finally, as China's economy and statistical system continue to evolve, indicators that do well historically might do less well going forward. For example, that government revenue and real estate investment do better after 2008Q1 than during the 2000Q1-2007Q4 period could reflect either changes in the composition of activity, changes in the quality of activity, or could be chance. Nevertheless, it is reassuring that our core set of indicators performs well across our two sample periods.

References

Bai, J., Perron, P., (1998), "Estimating and testing linear models with multiple structural changes," Econometrica 66, 47–78.

Bai, J., Perron, P., (2003), "Computation and analysis of multiple structural change models," Journal of Applied Econometrics 18, 1–22.

Bradsher, Keith (2012). "China Data Mask Depth of Slowdown, Executives Say," New York Times, June 23. http://www.nytimes.com/2012/06/23/business/global/chinese-data-said-to-be-manipulated-understating-its-slowdown.html?pagewanted=all

Clark, Hunter, Maxim Pinkovskiy, and Xavier Sala-i-Martin, 2017."China's GDP Growth May be Understated." NBER Working Papers 23323, National Bureau of Economic Research, Inc.

Dawson, John W., Joseph P. DeJuan, John J. Seater, and E. Frank Stephenson. "Economic Information versus Quality Variation in Cross-Country Data." The Canadian Journal of Economics / Revue Canadienne D'Economique 34, no. 4 (2001): 988-1009. http://www.jstor.org/stable/3131934.

Fernald, John, Israel Malkin, and Mark M. Spiegel, (2013), "On the Reliability of Chinese Output Figures," FRBSF Economic Letter, 2013-08, March 25.

Fernald, John & Edison, Hali, and Loungani, Prakash, (1999), "Was China the first domino? Assessing links between China and other Asian economies," Journal of International Money and Finance, Elsevier, vol. 18(4), pages 515-535, August.

Hansen, Peter Reinhard and Allan Timmermann, (2012), "Choice of Sample Split in Out-Of -Sample Forecast Evaluation," EUI Working Paper NO. ECO 2012/10.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, (2012), "Measuring Economic Growth from Outer Space," American Economic Review, 102(2), 994–1028.

Holz, Carsten A. (2008). "China's 2004 Economic Census and 2006 Benchmark Revision of GDP Statistics: More Questions than Answers?" The China Quarterly, March.

Holz, Carsten A. (2013). "Chinese Statistics: Output Data"

Holz, Carsten A. (2003). "'Fast, Clear and Accurate:' How Reliable Are Chinese Output and Economic Growth Statistics?" The China Quarterly, no. 173 (March 2003): 122-63.

Kelly, Bryan and Seth Pruitt, (2013), "Market Expectations in the Cross-Section of Present Values, *The Journal of Finance*, 68(5), 1721-1756.

Nakamura, Emi, Jon Steinsson, and Miao Liu, (2014), "Are Chinese Growth and Inflation Too Smooth?: Evidence from Engel Curves," mimeo, Columbia University, January.

Noble, Josh (2015). "Doubts rise over China's official GDP growth rate." *Financial Times,* September 16, 2015. http://www.ft.com/intl/cms/s/0/723a8d8e-5c53-11e5-9846-de406ccb37f2.html#axzz3lvULIfcE (accessed September 21, 2015).

Pinovsky and Sala-i-Martin (2016) QJE.

Sharma, Ruchir (2013). "China's Illusory Growth Numbers." Wall Street Journal, October 31, 2013.

Sinclair, Tara (2012). "Characteristics and Implications of Chinese Macroeconomic Data Revisions." Manuscript, George Washington University. http://www.gwu.edu/~iiep/assets/docs/papers/Sinclair_IIEPWP2012-09.pdf

Stock, James H., and Mark W. Watson. 2016. "Dynamic Factor Models, Factor Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In Handbook of Macroeconomics, Volume 2A, edited by John B. Taylor and Harald Uhlig. Amsterdam: North-Holland.

Wikileaks (2007). http://wikileaks.org/cable/2007/03/07BEIJING1760.html.

Wright, Jonathan H. 2017. "Optimal Seasonal Filtering." Manuscript, Johns Hopkins University.

Appendix: (Still in separate document for ease of simultaneous editing)

Appendix:

1. Data sources

The chart below shows the raw data we used in the paper. All data were accessed in May 2017, mainly from CEIC Asia database.

Series	Description	Source
Electricity	Electricity production, Billions of kilowatt hours	National Bureau of Statistics (CEIC series 3662501)
Rail	Railway freight traffic, millions of tons	China Railway Corporation, National Railway Administration (CEIC series 12915101)
Lending	Bank loans, billions of RMB	The People's Bank of China (CEIC series 7029101)
Property	Real estate investment (Residential bldgs.), millions of RMB	National Bureau of Statistics (CEIC series 3948701)
Air passengers	Air passenger traffic, millions of persons	Civil Aviation Administration of China (CEIC series 12916401)
Exports	Exports (FOB basis), millions of US dollars	General Administration of Customs (CEIC series 5823501)
Consumer index	Consumer Expectation Index	National Bureau of Statistics (CEIC series 5198601)
Floor space	Floor space started, thousands of square meters	National Bureau of Statistics (CEIC series 3963901)
Raw materials	Index of raw materials supply, derived from a survey of managers from 5000 companies. Respondents are asked for views on adequacy of supplies of raw materials.	The People's Bank of China (CEIC series 8003501)
Retail	Retail sales of consumer goods, billions of RMB	National Bureau of Statistics (CEIC series 5190001)
Industrial production	Value added of industry, YoY %	National Bureau of Statistics (CEIC series 3640701)
Highway	Freight carried, highway, Ton mn	China Economic Monitoring & Analysis Center, NBS (CEIC series 12915201)
Government revenue	Billions of RMB	Ministry of Finance (CEIC series 4331701)
FAI	Fixed asset investment, billions of RMB	National Bureau of Statistics (CEIC series 7872901)
GDP	Real GDP index, available as 4-quarter growth rates	National Bureau of Statistics (CEIC series 1692001)

Exchange rates between Yen, Euro, USD, and RMB		Bloomberg
Imports and exports between Japan and other countries	Thousands of Yen	Ministry of Finance (http://www.customs.go.jp/toukei/info/index_e.htm)
Imports and exports between European Union and other countries	Euros	Eurostat ("EU27 trade since 1988 by CN8" database)
Imports and exports between United States and other countries		Census Bureau via Haver Analytics
Other imports and exports between China	Direction of Trade Statistics	IMF

2. Calculation of export price index

We deflate trading-partner exports to China using a Chinese-specific deflator for U.S. exports. This deflator weights growth in overall U.S. agriculture and non-agriculture deflators by corresponding shares of U.S. exports to China and Hong Kong. Fernald, Malkiel, and Spiegel (2013) found that this “simple” deflator corresponds closely to a more sophisticated deflator for exports to China that uses detailed commodity-by-country data. (Those detailed data are available only after 2005:12.) The data sources are U.S. Census data on "Trade in Goods by NAICS-Commodity By Country," and "Export Price Indexes by NAICS." Data were accessed via Haver Analytics August 3, 2015.

3. Other adjustments

Monthly proxy series converted to quarterly via summing over the quarter.

Missing observations around Chinese New Year:

- ElectricityConsumption missing January, filled in with half of February cumulative value

- Retail missing January and February, 1st quarter percent change filled in with March-to-March 4-quarter change
- Rail shipments: Series has a break in level in January 2004. We adjusted the series by splicing. (Done in code d02_input_proxy_data.do).
- Li: We use the adjusted rail data rather than the raw data.

<u>TRADE</u>

Data on China and Hong Kong summed and used for Chinese import/export numbers All trade series converted to USD using bilateral exchange rate data (Bloomberg, “USDEUR Curncy” and “USDJPY Curncy” series). Thus:

- EU trade in millions USD = (EU trade in Euros)/(1000000*(Euro-USD exchange rate)
- Japanese trade in millions USD = (Japanese trade in thousands Yen)/(1000*Yen-USD exchange rate)

Table 1: Statistical Capacity and Correlation of Import and GDP Growth

	(1)	(2)	(3)
log(GDP per capita) 1990	0.073***		0.011
	(0.019)		(0.021)
Trade/GDP 1990	-0.059		-0.005
	(0.048)		(0.046)
Statistical capacity: B		-0.035	-0.029
		(0.099)	(0.100)
Statistical capacity: C		-0.340***	-0.321***
		(0.069)	(0.078)
Statistical capacity: D		-0.426***	-0.401***
		(0.078)	(0.092)
Constant	-0.095	0.788***	0.677**
	(0.160)	(0.064)	(0.217)
Observations	165	165	165
R-squared	0.082	0.217	0.218
Standard errors in parentheses	* p<0.05	** p<0.01	*** p<0.001

Notes: Dependent variable is the correlation of growth rates of real imports and real GDP in annual data from 1990 to 2016 (??) for a cross-section of countries. Real imports and real GDP are in constant national currencies. Statistical capacity is from the Penn World Tables version 6.1.

Table 2: Out-of-sample (OS) Results for Different In-Sample (IS) Periods

	2000 - 2014	2015 - 2016	2008 - 2014	2015 - 2016
Indicators	IS RMSE	OS RMSE	IS RMSE	OS RMSE
Electricity	0.66	0.56	0.66	0.59
Exports	0.69	0.82	0.84	0.80
Industrial production (IP)	0.72	0.59	0.79	0.74
Rail	0.85	0.69	0.63	1.22
Floor space	0.86	0.56	0.87	0.64
Government revenue	0.86	0.83	0.85	0.89
GDP	0.88	0.67	0.64	0.76
Property	0.88	0.39	0.73	0.40
Consumer index	0.92	0.58	0.82	0.69
Fixed Asset Investment (FAI)	1.04	0.75	1.28	1.11
Highway	1.04	0.84	1.25	0.89
Lending	1.04	0.81	1.27	1.04
Air passengers	1.04	0.82	1.27	1.10
Retail	1.04	0.85	1.26	0.94
Li Index	0.70	0.51	0.61	0.60
All Indicators	0.64	0.35	0.54	0.50

Notes: The numbers shown are the root-mean-squared errors (RMSEs) from predicting import growth using the indicator shown (along with China's real exchange rate). The left two columns show estimates when the in-sample (IS) period for estimating the regressions is from 2000Q1 through 2014Q4. The right two columns show results when the in-sample estimation period is 2008Q1-2014Q4. In both cases, the out-of-sample (OS) period for calculating RMSEs is 2015Q1 through 2016Q4. The OS results use the regression coefficients estimated in the corresponding in-sample period.

Table 3: Sequential Indicators and Average IS-OS Rank

NumVars	Variables	RMSE Avg
1	Electricity	0.61
2	Electricity Property	0.53
3	Electricity Property IP	0.53
4	Electricity Property IP FloorSpace	0.54
5	Electricity Property IP FloorSpace ConsumerIndex	0.52
6	Electricity Property IP FloorSpace ConsumerIndex Exports	0.52
7	Electricity Property IP FloorSpace ConsumerIndex Exports Rail	0.49
8	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP	0.48
9	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP GovtRev	0.46
10	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP GovtRev FAI	0.47
11	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP GovtRev FAI Lending	0.48
12	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP GovtRev FAI Lending AirPassengers	0.48
13	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP GovtRev FAI Lending AirPassengers Highway	0.50
14	Electricity Property IP FloorSpace ConsumerIndex Exports Rail GDP GovtRev FAI Lending AirPassengers Highway Retail	0.50

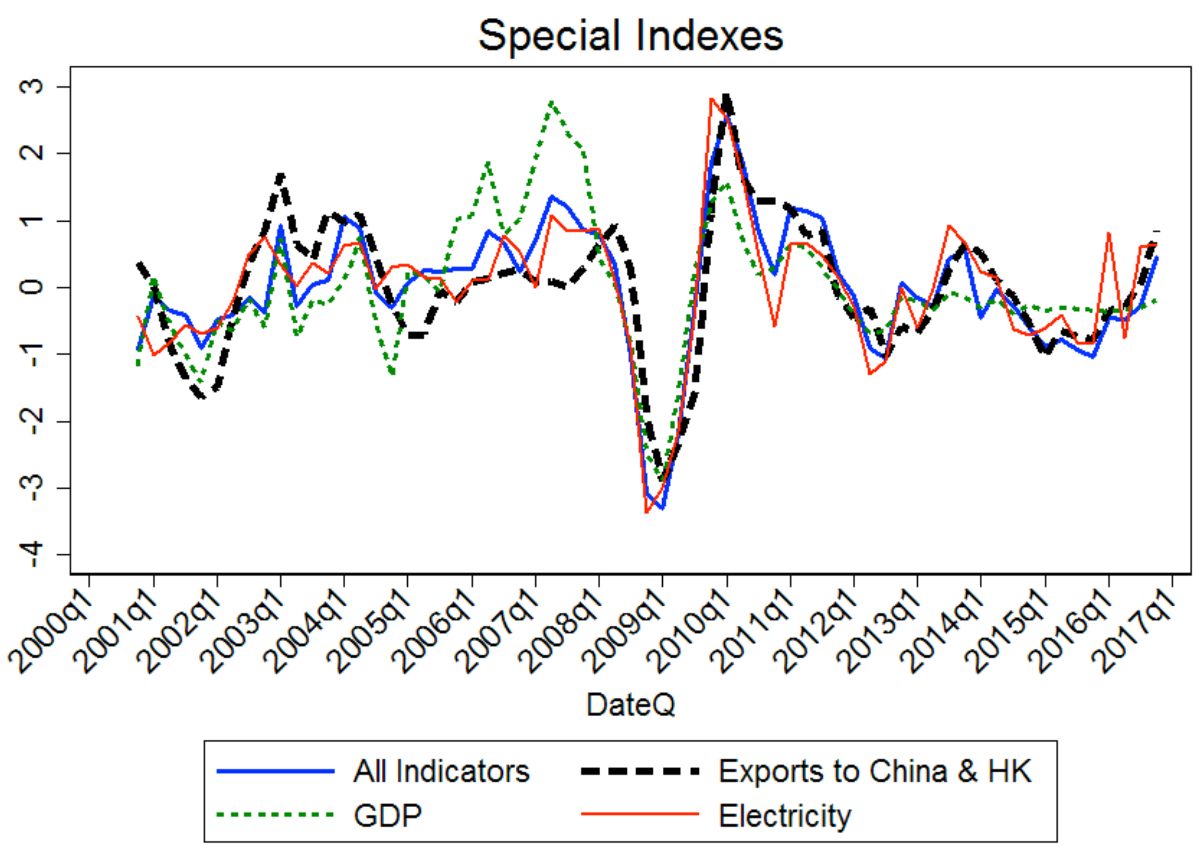
Notes: The first row shows the indicator with the lowest root-mean-squared error (RMSE), averaged over the in-sample (IS) and out-of-sample (OS) periods. The subsequent rows proceed sequentially by adding the indicator with the next lowest average RMSE to the list of indicators on the preceding line. IS estimation is from 2000Q1-2014Q4. OS RMSEs use the estimated regression coefficients from the IS estimation.

Table 4: Average IS-OS Rank

NumVars	Variables	RMSE Avg
1	Electricity	0.61
2	GovtRev Rail	0.50
3	GovtRev IP Rail	0.44
4	Electricity Exports GovtRev Rail	0.42
5	Electricity Exports GovtRev Rail Retail	0.42
6	AirPassengers Electricity Exports GovtRev Rail Retail	0.43
7	AirPassengers Electricity Exports FloorSpace GovtRev Rail Retail	0.43
8	AirPassengers ConsumerIndex Electricity FAI FloorSpace GovtRev Rail Retail	0.44
9	AirPassengers Electricity Exports GDP GovtRev Lending Property Rail Retail	0.44
10	AirPassengers ConsumerIndex Electricity FloorSpace GDP GovtRev Lending Property Rail Retail	0.45
11	AirPassengers ConsumerIndex Electricity Exports FloorSpace GDP GovtRev Lending Property Rail Retail	0.45
12	AirPassengers ConsumerIndex Electricity Exports FAI FloorSpace GDP GovtRev Lending Property Rail Retail	0.46
13	AirPassengers ConsumerIndex Electricity Exports FAI FloorSpace GDP GovtRev IP Lending Property Rail Retail	0.48
14	AirPassengers ConsumerIndex Electricity Exports FAI FloorSpace GDP GovtRev Highway IP Lending Property Rail Retail	0.50

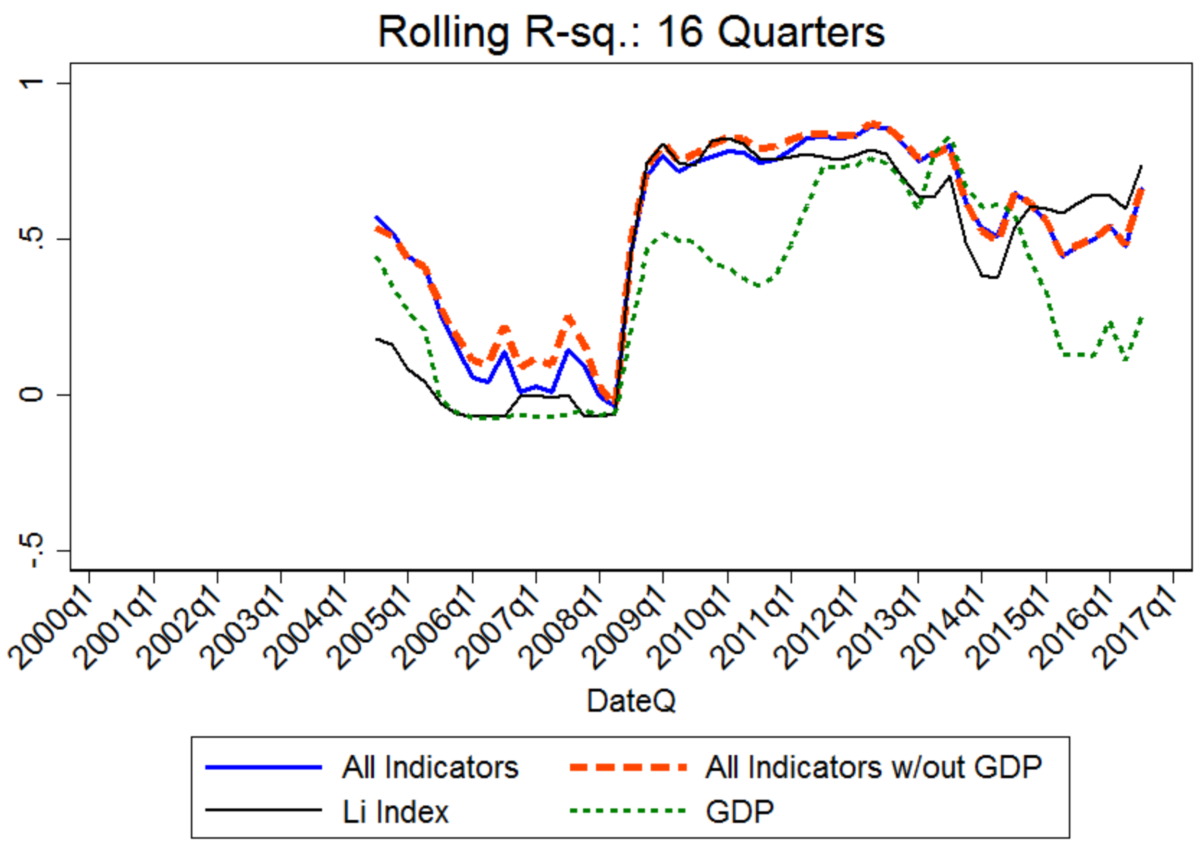
Notes: For each number of indicators (from 1 through 14), the rows show the combination of all possible indicators with the lowest root-mean-squared error (RMSE), averaged over the in-sample (IS) and out-of-sample (OS) periods. IS estimation is from 2000Q1-2014Q4. OS RMSEs use the estimated regression coefficients from the IS estimation.

Figure 1: Indicators of economic activity in China



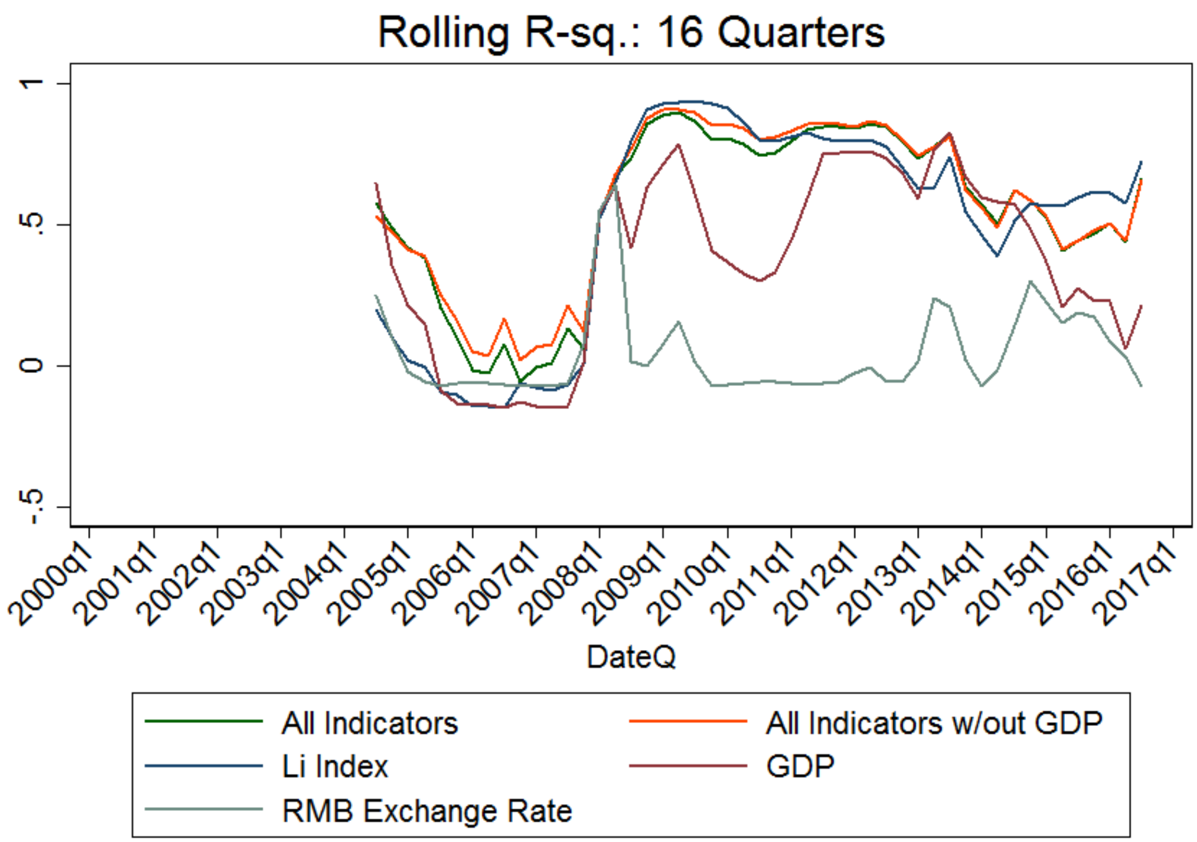
Note: Normalized GDP and "BROAD" exports to China 2000Q4-2014Q4. "All indicators" series is normalized first principal component of all 14 activity indicators. See text for details.

Figure 2: Rolling R-Squareds (NO EXCH RATE—TO DROP)



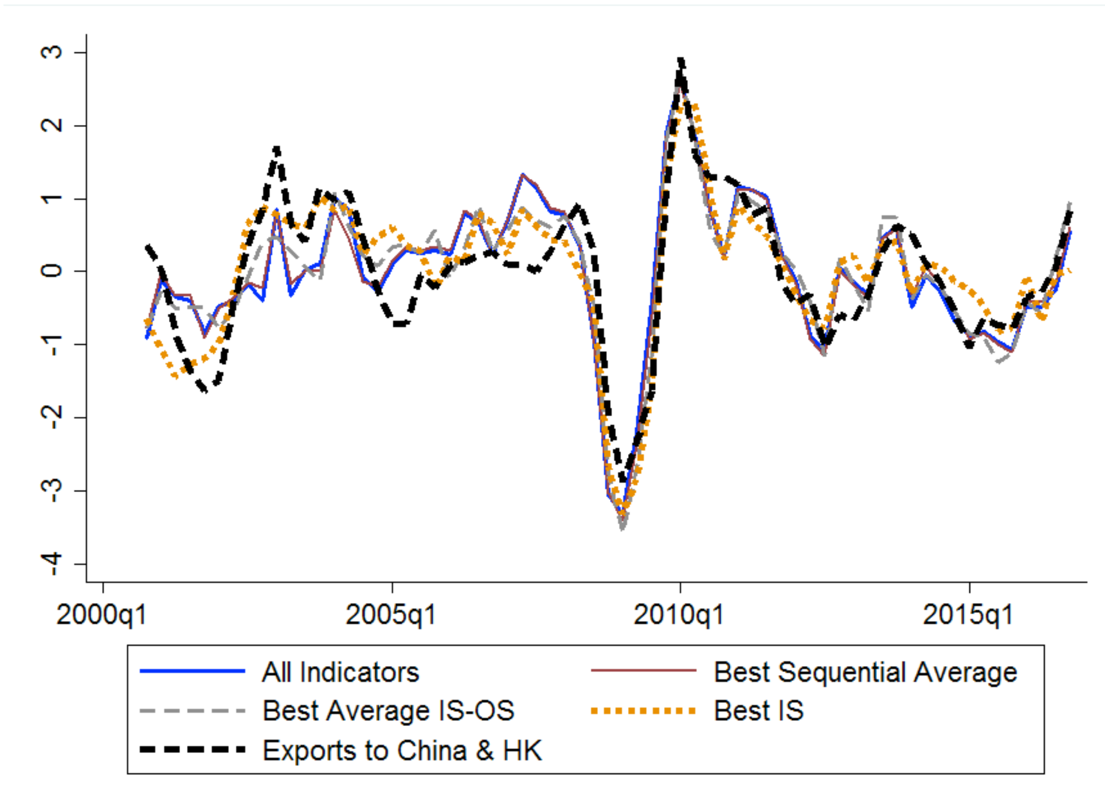
Note: R-squared values for rolling samples over 16-quarters, regressing Chinese imports on first principal components of all 10 indicators, GDP, and Li indicators, respectively. Chinese nominal exchange rate changes and constant term included in all specifications.
:

Figure 2: Rolling R-Squareds (with RMB XR)



Note: R-squared values for rolling samples over 16-quarters, regressing Chinese imports on first principal components of all 10 indicators, GDP, and Li indicators, respectively. Chinese nominal exchange rate changes and constant term included in all specifications.
:

Figure 3: Indicators and Imports



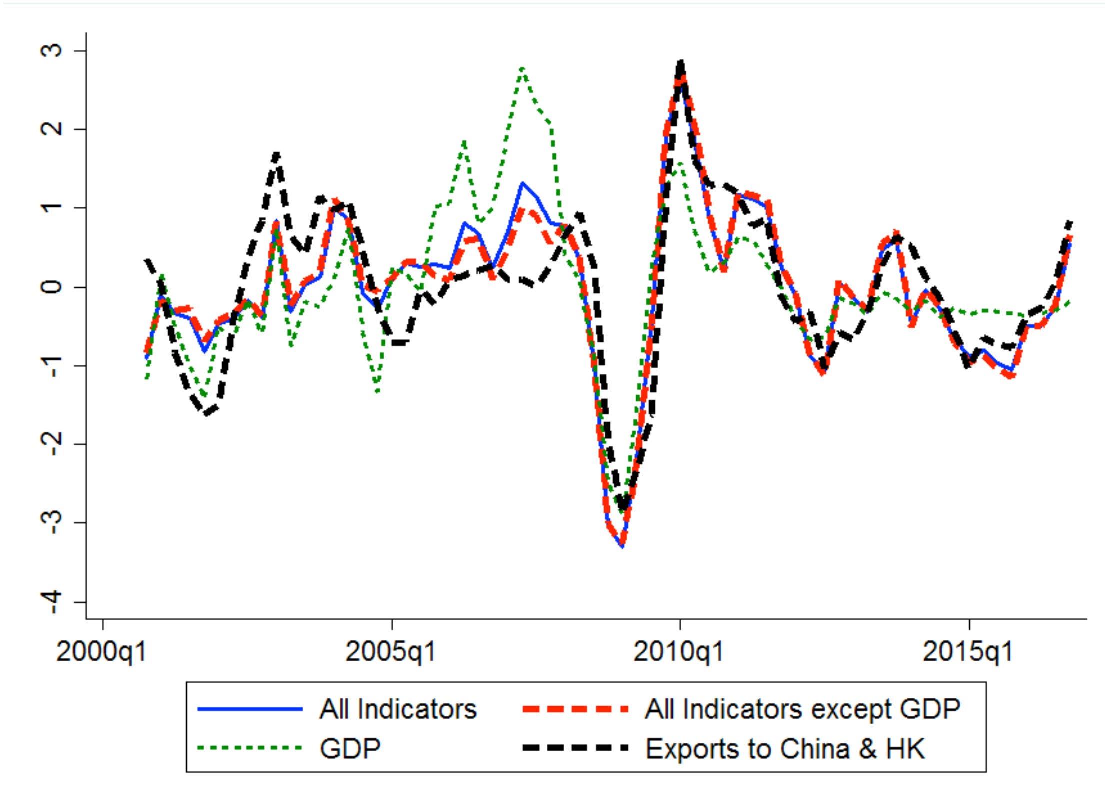
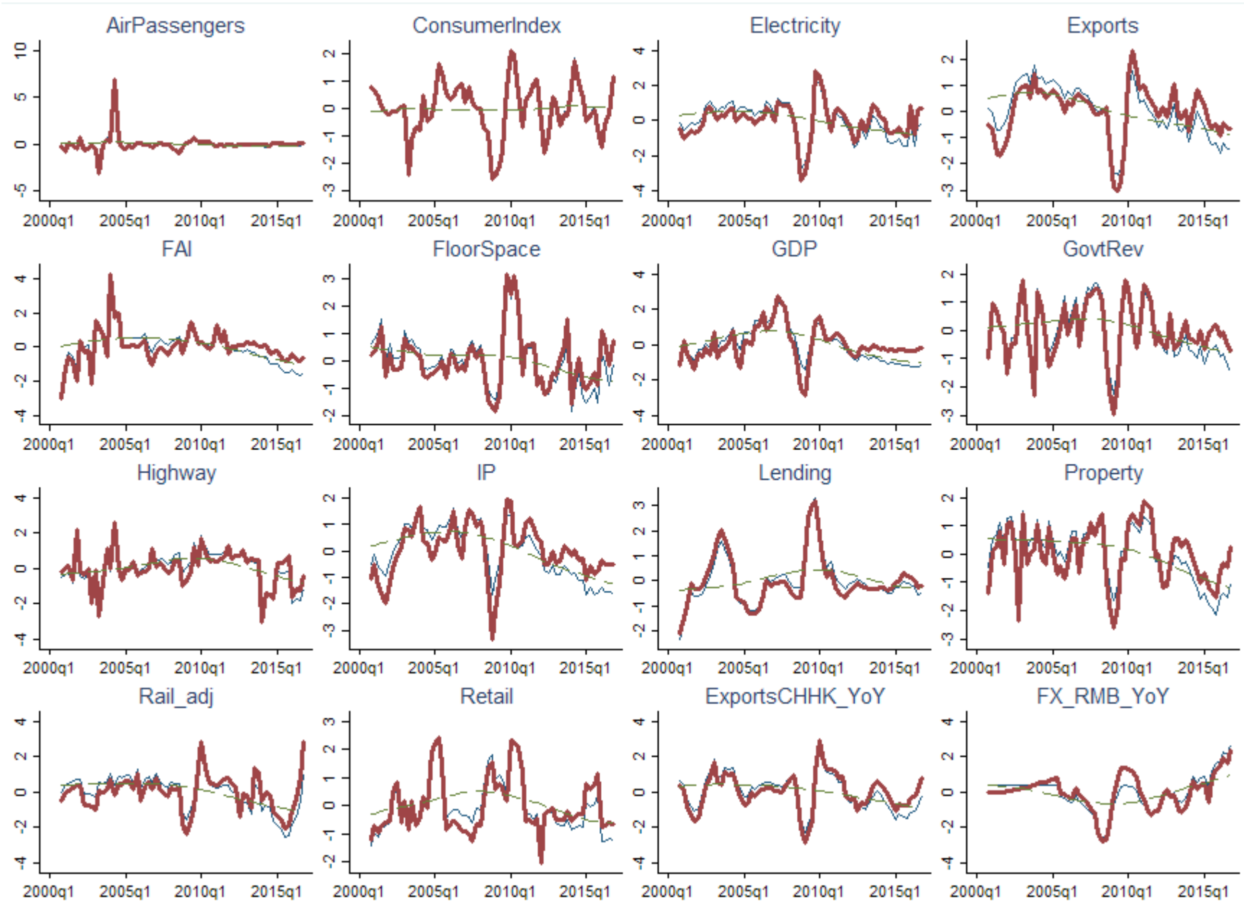

3
2
1
0
-1
-2
-3
-4
2000q1
2005q1
2010q1
2015q1
All Indicators
All Indicators except GDP
GDP
Exports to China & HK

Figure A1: Individual Indicators



Note: Blue is normal; red is filtered normal; green is trend.

Figure A2: Exports to China Versus China's Imports