Nowcasting China's PPI inflation using low-frequency and mixed-frequency dynamic factor models*†‡

Run Liang §
Peking University

Fa Wang¶
Peking University

Jiawen Xu‖ **
Shanghai University of Finance and Economics
Key Laboratory of Mathematical Economics (SUFE)

Abstract

We construct nowcasts and forecasts of China's PPI inflation using a large panel of data series with different frequencies (monthly, ten-day, weekly, daily). Mixed-frequency data are incorporated into the dynamic factor model via two approaches: one is to convert high-frequency data to low-frequency data, monthly in our example, and apply the expectation maximization (EM) algorithm for estimation; the other is to treat low-frequency data as high-frequency (daily) data with missing observations in a specific pattern and apply Bańbura and Modugno (2014) to estimate the dynamic factor model with missing observations. We compare the forecast accuracy of these two approaches with some other alternative forecasting models, such as a random walk, a univariate autoregressive model and a dynamic factor model with only monthly data. Our empirical results show that the first approach outperforms other models in most cases and horizons. Models utilizing high-frequency data generally perform better than those that do not. As high-frequency information flows in, forecast accuracy improves substantially.

Key words: Factor Models; Mixed Frequencies; PPI inflation; Nowcasting
JEL Classifications: C32, C53, E31, E37

---

*This research is supported by the National Natural Science Foundation of China (Grant no. 71873085; 71773069; 72003117) and High-performance Computing Platform of Peking University.

†The authors would like to thank Michele Modugno for sharing his dynamic factor model code.

‡The authors are responsible for all errors.

§Peking University HSBC Business School, University Town, Nanshan District, Shenzhen, 518055, China (liangrun@pku.edu.cn).

¶School of Economics, Peking University, 5 Yiheyuan Road, Beijing, 100871, China (fa.wang@pku.edu.cn).

‖(corresponding author)Institute of Advanced Research, Shanghai University of Finance and Economics, 777 Guoding Rd., Shanghai China 200433 (xu.jiawen@mail.shufe.edu.cn).

**Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai, 200433, China

1 Introduction

Forecasting short-term inflation is of primary importance to central bank policy makers in most countries. Two often-used measures of inflation, CPI and PPI, represent the finished goods sector and the intermediate goods sector, respectively. As can be observed in the data, the correlation between the two has declined substantially since the start of this century; see Wei and Xie (2018), which leads us to rethink traditional monetary policy rules and determine the underlying shocks that may affect price stability. We have seen some theoretical papers discussing the preferred target in an optimal monetary policy reaction function with different settings of the economy and assumptions about the production model, i.e., Huang and Liu (2005); Gali and Monacelli (2005), De Paoli (2009), and Wei and Xie (2020), but fewer empirical works focus on forecasting PPI inflation.

Thanks to globalization and the lengthening of global production chains, China has gradually become the world's manufacturing powerhouse since 2001. Macroeconomists or policy makers should be aware of the consequences of ignoring fluctuations in the prices of intermediate goods in such a large open economy with a multistage production network. The empirical analysis in this paper focuses on China's PPI inflation. More accurate and timely forecast revisions can be made by including more relevant data groups. The relevant data groups include variables that are highly correlated with PPI inflation and released earlier than the relevant inflation releases. More important, data with different frequencies (i.e., daily, weekly, ten-day, monthly) are utilized, and information flows in a sequential manner, making continuous revisions feasible.

The main challenges of the empirical analysis are due to the specific characteristics of the data set. We attempt to extract predictive content from a large unbalanced panel with different frequencies and different degrees of timeliness (release lags). For example, 10-day data on the producer goods market price have only been available since 2014, but they may contain reliable and timely information on the evolution of manufacturing materials. The econometric framework that we adopt in this paper is the factor model. The works most related to our paper are Lenza and Warmedinger (2011) and Modugno (2013), which they handle mixed-frequency data in different ways. The former paper first converts higher frequencies into low frequencies and then applies a dynamic factor model to compute forecasts of the target low-frequency variable, while the latter adopts the estimation methodology proposed in Bańbura and Modugno (2014) and models data as a trading day frequency factor model with missing observations cast in a state space representation. Both approaches have advantages and disadvantages. The estimation of the first approach is standard and parsimonious. However, the conversion of high-frequency data by simply taking averages may result in some loss of information content for high-frequency variables. The second approach can merge all data within a single, unified, coherent framework and, as a result, estimate many more coefficients. One possible reason that may lead to underperformance is that the assumption of treating low-frequency variables as high-frequency variables with missing observations sometimes captures and amplifies excessive high-frequency fluctuations, which are unobserved or smoothed out in the low-frequency target variables, but miss informative correlations in the low-frequency components. Moreover, treating monthly inflation as a snapshot in the middle of the month may cause inaccuracy in forecasting, especially when the monthly average differs substantially from the snapshot at the middle of the month. Surprisingly, no comparative studies have been conducted on this subject. We fill this gap by comparing the forecasting accuracy of these two approaches. Another relevant paper by Knotek and Zaman (2017) argues that when using real-time and subjectively chosen variables, a simple parsimonious predictive regression can outperform models considering large data sets and mixed-frequency data in nowcasting U.S. CPI inflation.

The empirical evidence shows that when forecasting China's PPI inflation, the low-frequency dynamic factor model (DFM), which converts higher-frequency data into lower-frequency data, outperforms the mixed-frequency DFM at all horizons. Forecasting accuracy will also increase when one includes more types of relevant data groups, such as the producer goods market price released every 10 days. Compared with models using only monthly data, simple random walk, and autoregressive models, models utilizing high-frequency data have significantly stronger predictive power in most cases. High-frequency data are necessary to produce more accurate inflation forecasts. Moreover, the empirical evidence evaluates the subcomponents of inflation, suggesting that the forecast improvement is due to the more accurate forecasts of the mining and quarrying, raw material, and manufacturing components. Robustness checks in different subsamples confirm our baseline results. We also discover a pattern of increasing accuracy when forecasting through the timeline of the data release.

The remainder of the paper is organized as follows: Section 2 describes different data groups, and Section 3 introduces two factor model-based approaches. In Section 4, we present the comparisons of forecasting accuracy, and in Section 5, we assess the importance of timeliness in higher frequency data. Finally, Section 6 concludes the paper.

2 Data

High-frequency data releases in China began recently, and some series have been discontinued in recent years. How to ensure high-quality data collection has become a common issue in most developing countries. In this paper, we collect our data according to three criteria: 1) The series should have a sample period of at least five years. 2) The price series should be a good representative of production materials. 3) The price series should reflect price movements at the country level. A list of all of the data is provided in Table A.1. Our main data source is CEIC's China Premium Database[1], which compiles China's official macroeconomic time series, and the WIND financial Terminal[2].

More precisely, the first group of variables in our data set includes the market price of raw materials (RMP). There are 11 series included in this group. These series are sampled at a daily frequency, collected from various sources and published every trading day. This group of variables may capture pricing information in the early stage of the pricing chain, therefore indicating fundamental price developments. Figure 1 shows that there is indeed a high degree of correlation between the average month-on-month growth rate of the RMP series and the month-on-month growth rate of total PPI, and RMP leads PPI at some turning point. For example, during the financial crisis in 2008, the RMP growth rate reached its minimum one month earlier than PPI growth. The lowest PPI growth rate occurred in November 2008, which was -4.6% after standardization and centering on zero, while the lowest RMP growth rate occurred in October 2008, which was -4.8% after standardization and centering on zero, just one month before the former.

The second group contains the China Commodity Price Index (CCPI). There are 10 indexes included in the CCPI, which covers aggregate, energy, steel, mineral production, nonferrous metals, rubber, agricultural products, livestock, oil & oil seed, and sugar. These indexes are constructed by the Ministry of Commerce of the People's Republic of China based on the weekly commodity spot price database of the China International E-Commerce Network. These data are sampled at a weekly frequency and are published weekly on the Wednesday of the week following the reference week. Figure 2 reveals that the average month-on-month growth rate of the CCPI series is highly correlated with the month-on-month growth rate of total PPI. We observe similar patterns as in figure 1: the CCPI growth rate leads the PPI growth rate by one month during economic downturns.

[1] CEIC website: https://www.ceicdata.com/en

[2] WIND website: https://www.wind.com.cn/en/wft.html

The third group includes the producer goods market price (PGMP) released by the National Bureau of Statistics of China. They are sampled at a 10-day frequency and are published every 10 days on the fourth day of the next 10-day period. There are 47 series included in the PGMP, which covers various products of industrial goods and agricultural products, which may provide more information on total PPI. However, this group of data began to be released later than the others, in January 2014, so we only include these data for subsample analysis. Figure 3 shows that the average month-on-month growth rate of PGMP is highly correlated with the month-on-month growth rate of total PPI. Again, the PGMP growth rate also leads the PPI growth rate. For example, the highest growth rate of PGMP was achieved in November 2016, while the highest growth rate of PPI was achieved one month later in December 2016.

The fourth group includes monthly surveys of the PMI purchasing price index and becomes available on the last day of the reference month, i.e., 15 days before the PPI release. The PMI purchasing price index directly reveals the change in the purchasing price of raw materials and intermediate goods and may provide more precise information about the month-on-month price change of producer goods than the aggregate PMI. It contains information on the last stage of the pricing chain. Figure 4 shows that the PMI purchasing price index comoves to a large extent with the month-on-month growth rate of PPI.

The fifth group includes financial data at a daily frequency. In particular, it includes the CSI 300 stock index (CSI 300), the 7-day interbank offering rate (IOR 7d), the three-month Treasury bill rate, the ten-year Treasury bond rate, and the central parity rate of the RMB against the USD (USDCNY). Such financial data may incorporate expectations on the future development of prices and are available on a daily basis. The forward-looking nature of financial asset prices is relevant for macroeconomic forecasting; see Monteforte and Moretti (2013) and Andreou et al. (2013) for reference.

Fig. 5-9 plots the chosen financial variables (the month-on-month growth rate of CSI 300 and the original level of other variables) vis-à-vis the month-on-month growth rate of total PPI. Figure 5 shows that the month-on-month growth rate of CSI 300 fluctuates more frequently than the month-on-month growth rate of total PPI and has leading indicator properties vis-à-vis total PPI inflation. Concerning the interest rates, shown in Figs. 6-9, the very short-term interest rate, i.e., IOR 7d, fluctuates more than the longer-term rates, i.e., TB 3 M and TB 10Y. All the interest rates seem to reflect the low-frequency component of total PPI inflation. As shown in Figure 9, the exchange rate seems to reflect even longer trends of total PPI inflation. Overall, the correlation of the financial data with total PPI inflation does not appear as strong as the high-frequency price series does. However, we still include these data to obtain future information.

Finally, despite the total PPI, we also include seven components, i.e., mining and quarrying, raw material, manufacturing, food, clothing, daily sundries, and durables. The first three components are categorized as producer goods, with aggregate weights of approximately 75%, and the latter four components are categorized as consumer goods, with aggregate weights of approximately 25%. By assessing the effects of the remaining groups of variables (RMP, CCPI, PGMP, PMI, and financial data) on the forecast of each individual PPI component, we shed light on the underlying forces driving the improvement in the forecast accuracy of total PPI.

To enable a better and easier understanding of the timeliness of each data release for the PPI data, we present a timeline of the various groups of variables employed for a given month as they appear. Overall, the following logic applies: the first release of PPI data in China is published approximately 15 days after the reference month. Before that day, however, the remaining variables in our data set, which already contain preliminary evidence/information about the PPI of the current month, have already been released. Figure 10 provides an illustrative example of the timeliness of the data employed to forecast PPI inflation in July 2017, our reference period in this example. Note that the PPI data for July 2017 (i.e., the target forecast) are released on 15 August and therefore are not presented in our July timeline. In terms of July PPI data releases, the only available information is the PPI for June, which is released on 15 July (see the upper part of Fig. 10). Note that the actual date of PPI release varies from the 9th to 15th of each month; thus, we set the release date of monthly PPI to the 15th for simplicity.

Focusing on the four variable groups in our data set, it becomes clear that they are released in a much more timely manner than the PPI data. We can see from the lower part of Fig. 10 that several sets of high-frequency data are released gradually, containing information about June and (mainly) July. Specifically, the first high-frequency data released in July 2017 are the PGMP of June 21-30 on July 4. Then, on July 5, the weekly CCPI data from June 26 to July 2 are released. The remaining weekly CCPI data are updated every week (July 12, 19, and 26). In a parallel fashion, the July 1-10 and 11-20 PGMP data are released on July 14 and 24, respectively. Now, let us consider the upper part of Fig. 10, which shows the release date of the monthly series published in July 2017. The July PMI surveys are released on the last day of the month, July 31, and contain information on the current month. Moreover, daily RMP data and financial data for the current day are also available on each trading day (not shown in the timeline).

3 Methodology

3.1 Low-Frequency Dynamic Factor Model

The low-frequency DFM is given as follows:

$$y_t = \Lambda f_t + \varepsilon_t,\ \ \varepsilon_t \sim i.i.d.N\left(0, R\right) \tag{1}$$

where $y_t = \left(y_{1t}, y_{2t}, \ldots, y_{nt}\right)'$ is an $n \times 1$ vector of observations at month t (t=*1,..., T*); here, y_t is the month-on-month growth rate of monthly series. Λ is an $n \times r$ matrix of factor loadings, $\varepsilon_t = \left(\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt}\right)'$ is an $n \times 1$ vector of series specific idiosyncratic components, and f_t is an $r \times 1$ vector of unobserved common factors which follows a VAR process:

$$f_t = Af_{t-1} + u_t, \quad u_t \sim i.i.d.N\left(0, Q\right) \tag{2}$$

where A is an $r \times r$ matrix and u_t is an $r \times 1$ vector.

To address the mixed frequency issue and utilize high-frequency information, we convert these high-frequency series into monthly series by taking the monthly average as in Giannone et al. (2008) when a full month of observations is available. When only some of the observations are available, we assume the series to follow a random walk at the higher frequency to supplement the unavailable data in the current month and then take an average, as in Knotek and Zaman (2017). Note that there might be no observation for some series in the current month, which will lead to missing data issues. Moreover, some series released recently may contain missing values at the beginning. Bańbura and Modugno (2014) shows how to estimate the parameters when there are arbitrary patterns of missing data.

3.2 Mixed-Frequency Dynamic Factor Model

According to the different sampling frequencies, we divide the data into four groups and define them as $Z_s^{(m)}$, $Z_s^{(10d)}$, $Z_s^{(w)}$, and $Z_s^{(d)}$.

- $Z_s^{(m)}$ denotes the logarithm of the monthly series Z in month m and on day s. There are k_m trading days between any two consecutive releases.
- $Z_s^{(10d)}$ denotes the logarithm of the 10-day series Z in a 10-day period $10d$ and on day s. There are k_{10d} trading days between any two consecutive releases.

- $Z_s^{(w)}$ denotes the logarithm of the weekly series Z in week w and on day s. There are k_w trading days between any two consecutive releases.

- $Z_s^{(d)}$ denotes the logarithm of the daily series Z observed on each day s.

Then, $z_s^{(m)}$,$z_s^{(10d)}$,$z_s^{(w)}$, and $z_s^{(d)}$ are defined as the monthly, 10-day, weekly, and daily growth rate accordingly, i.e., $z_s^{(m)} = (Z_s^{(m)} - Z_{s-k_m}^{(m)}) \times 100$, $z_s^{(10d)} = (Z_s^{(10d)} - Z_{s-k_{10d}}^{(10d)}) \times 100$, $z_s^{(w)} = (Z_s^{(w)} - Z_{s-k_w}^{(w)}) \times 100$, $z_s^{(d)} = (Z_s^{(d)} - Z_{s-1}^{(d)}) \times 100$.

As in Modugno (2013), when there are three types of data frequencies, the mixed-frequency DFM is written as follows:

$$\begin{bmatrix} z_s^{(m)} \\ z_s^{(w)} \\ z_s^{(d)} \end{bmatrix} = \begin{bmatrix} C_m & 0 & 0 \\ 0 & C_w & 0 \\ 0 & 0 & C_d \end{bmatrix} \begin{bmatrix} g_s^{(m)} \\ g_s^{(w)} \\ g_s^{(d)} \end{bmatrix} + \begin{bmatrix} \varrho_s^{(m)} \\ \varrho_s^{(w)} \\ \varrho_s^{(d)} \end{bmatrix} \tag{3}$$

where the transition equation becomes:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} g_s^{(m)} \\ g_s^{(w)} \\ g_s^{(d)} \end{bmatrix} = \begin{bmatrix} \Xi_s^{(m)} & 0 & 0 \\ 0 & \Xi_s^{(w)} & 0 \\ 0 & 0 & A \end{bmatrix} \begin{bmatrix} g_{s-1}^{(m)} \\ g_{s-1}^{(w)} \\ g_{s-1}^{(d)} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \nu_s^{(d)} \end{bmatrix} \tag{4}$$

When there are four types of data frequencies, the observation equation becomes:

$$\begin{bmatrix} z_s^{(m)} \\ z_s^{(10d)} \\ z_s^{(w)} \\ z_s^{(d)} \end{bmatrix} = \begin{bmatrix} C_m & 0 & 0 & 0 \\ 0 & C_{10d} & 0 & 0 \\ 0 & 0 & C_w & 0 \\ 0 & 0 & 0 & C_d \end{bmatrix} \begin{bmatrix} g_s^{(m)} \\ g_s^{(10d)} \\ g_s^{(w)} \\ g_s^{(d)} \end{bmatrix} + \begin{bmatrix} \varrho_s^{(m)} \\ \varrho_s^{(10d)} \\ \varrho_s^{(w)} \\ \varrho_s^{(d)} \end{bmatrix} \tag{5}$$

where the transition equation becomes:

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} g_s^{(m)} \\ g_s^{(10d)} \\ g_s^{(w)} \\ g_s^{(d)} \end{bmatrix} = \begin{bmatrix} \Xi_s^{(m)} & 0 & 0 & 0 \\ 0 & \Xi_s^{(10d)} & 0 & 0 \\ 0 & 0 & \Xi_s^{(w)} & 0 \\ 0 & 0 & 0 & A \end{bmatrix} \begin{bmatrix} g_{s-1}^{(m)} \\ g_{s-1}^{(10d)} \\ g_{s-1}^{(w)} \\ g_{s-1}^{(d)} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \nu_s^{(d)} \end{bmatrix} \tag{6}$$

C_m, C_{10d}, C_w and C_d are the loadings for monthly, 10-day, weekly and daily variables, respectively; $g_s^{(m)}$, $g_s^{(10d)}$, $g_s^{(w)}$ and $g_s^{(d)}$ are the monthly, 10-day, weekly and daily factors sampled at the daily frequency; $\Xi_s^{(m)}$ is a time-varying coefficient that is equal to zero the day after each release of the monthly data and one otherwise; $\Xi_s^{(10d)}$ is equal to zero the day after each release of the 10-day data and one otherwise; $\Xi_s^{(w)}$ is equal to zero the day after each release of weekly data and one otherwise; and A is the matrix of the autoregressive coefficients for the daily factors. Once the model has been written in the state space form, it is straightforward to apply the methodology proposed by Bańbura and Modugno (2014).

According to the National Bureau of Statistics of China, PPI is constructed using the monthly average prices[3] of goods included in the PPI basket. To use the period mean data in the mixed-frequency model, we assume that the period mean is approximately the snapshot level observed in the middle of the period. In particular, when estimating the mixed-frequency factor model, we assume that $Z_s^{(m)}$ is a snapshot of prices observed in the middle of the month. Similarly, we treat the 10-day data as a snapshot of prices released on the 5th, 15th, and 25th days of the month. The weekly data are a snapshot of prices released on Wednesday every week. If the snapshot price on the release day turns out to be a fair approximation of the average prices in a certain period, e.g., a week, 10 days, a month, the mixed-frequency dynamic factor model would generate good forecasts. In contrast, if the snapshot price deviates considerably from the average, we would expect deterioration in forecasting performance, which we will show in detail in the empirical section.

4 Forecast Exercise: Design and Results

4.1 Forecast exercise design

To check whether high-frequency data can improve the PPI forecast accuracy and whether the transformation of high-frequency data to low-frequency data can perform better than the mixed-frequency model, we compare the following models:

- a low-frequency dynamic factor model that transforms all data into monthly frequency by taking monthly averages, estimated at the monthly frequency, by assuming random walk of the high-frequency series in the current month, as explained in Section 3.1. In later exercises, we transform 3 types of data (i.e., daily, weekly, and monthly) and 4

[3] The monthly average price is calculated as the simple average of the prices reported by individual enterprises around the 5th and 20th of the reference month.

types of data (i.e., daily, weekly, 10-day, and monthly) into monthly series and label them LF3 and LF4, respectively. [4]

- a mixed-frequency dynamic factor model that leaves all data frequencies unchanged, estimated at the trading day frequency, as explained in Section 3.2. In later exercises, we estimate mixed-frequency DFM including 3 types of data (labeled MF3) and 4 types of data (labeled MF4).[5]

- a dynamic factor model that includes only monthly variables (labeled Mon), i.e., PPI indexes and PMI surveys, and estimated at the monthly frequency. [6]

- a naïve random walk model (RW).

- a univariate autoregressive model with lag order one (AR(1)).

We adopt a recursive estimation scheme that, for the first evaluation, covers the period from January 2006 to January 2010. The out-of-sample forecasting evaluation spans the period from January 2011 to December 2020. Since the 10-day data are only available from January 2014, the evaluation sample of models including these data, i.e., LF4 and MF4, spans the period from January 2017 to December 2020. We calculate the out-of-sample forecasts for 0 (nowcast) to 12 horizons and report the root mean squared forecast errors (RMSFEs) for 0, 1, 3, 6, and 12 step-ahead results. Note that in all the above models, we forecast the month-on-month PPI growth rates, and then we compute the implied year-on-year growth rates.[7]

The data under analysis are never revised, which implies that, taking into account the data availability at each point in time at which we produce forecasts, we perform a real-time forecasting exercise.

Our model can be evaluated on any given day of the month. In this section, we show the results obtained on the 15th of each month, when PPI data for the previous month are

[4] Different parameters can be specified in terms of the number of factors and lag orders. We calculate forecasts from 24 factor models with 1 to 4 factors and 1 to 6 lags.

[5] We calculate forecasts from 160 factor models with 1 to 4 factors and 1 to 40 lags.

[6] The same combinations of factors and lag orders are calculated as in the low-frequency DFM model, i.e., 1 to 4 factors and 1 to 6 lags.

[7] Some papers assume a random walk or an autoregressive model in year-on-year growth rates rather than month-on-month growth rates to avoid seasonality; see Modugno (2013) and Knotek and Zaman (2017) for US CPI nowcasts. In China's PPI series, we do not observe strong seasonality. We also calculated the RMSFEs for the models using year-on-year growth rates. Most of the results are worse than those using month-on-month growth rates. Therefore, we continue to use to month-on-month growth rates for all the models.

released. As mentioned in the data description, the assumption of the PPI release date is merely for simplicity and has little effect on our main results. At that point in time, we already have some high-frequency data that contain information about the current month. For example, in July 2017 (see Figure 10), we already have two releases of CCPI data and one release of PGMP data containing information about July, available on July 5, 12, and 14, respectively. We also have 14 days of RMP data and financial data on that day. However, we do not have any PPI data for the current month, and the information available at a monthly frequency is the PPI and PMI survey data releases for the previous month (June). We choose this time point because the least high-frequency information is available at this point, so if the models containing high-frequency data can perform better than other models, it may be reasonable that they can outperform the others at another point in time. In Section 5, we show the evolution of the forecast accuracy obtained when evaluating the models at different points in time.

4.2 Baseline Results

This section presents the results of the alternative models described in Section 4.1. The results are presented for the best nowcasting models ex post that generate the smallest RMSFEs at horizon 0 from different combinations of factors and lag order choices (labeled LF3 BXP, MF3 BXP, and Mon BXP). However, to mimic a proper out-of-sample forecast exercise, we also present the arithmetic average of the RMSFEs produced by factor models that are characterized by different parameterizations (i.e., LF3 Avg). The LF3 model uses 3 types of data, i.e., daily, weekly and monthly, with a transformation to monthly frequency by taking the monthly average. The MF3 model uses the same data set without frequency transformation. The Mon model uses only monthly data. Moreover, we also report the results obtained with two univariate models, namely, a random walk (RW) and an AR(1) model.

The results for the full sample are reported in Table 1. This table compares the RMSFEs of the best model ex post of LF3 (LF3 BXP), MF3 (MF3 BXP), Mon (Mon BXP), and two univariate models RW and AR(1). The results show that the RMSFEs of our benchmark model, i.e., LF3 BXP, are smaller than those of the other models at all horizons for total PPI inflation and its major components.

First, we compare forecasting performance between the two factor models introduced in the previous section. By transforming the high-frequency data to low-frequency data, the low-frequency model performs much better than the mixed-frequency model, especially for nowcasts and short-term forecasts. For the nowcast (horizon 0) of total PPI, the RMSFE of LF3 BXP is 0.25, which is the smallest among all candidate models and is significantly smaller than that of MF3 BXP. The RMSFEs of LF3 BXP at horizons 1 and 12 are also significantly smaller than those of the mixed-frequency model. From the results of PPI subcomponents, we can see that most of the improvements are achieved in forecasting the mining and quarrying, raw material, and manufacturing components, which account for most variations in total PPI. The RMSFE for the nowcast of the mining and quarrying component is 1.95 for model MF3 BXP, while it is 1.40 for model LF3 BXP. For the raw material component, the RMSFE produced by model MF3 BXP at horizon 0 is 0.79, while LF3 BXP produces an RMSFE of 0.61. For the manufacturing component, the RMSFE of MF3 BXP at horizon 0 is 0.27, while it is 0.23 for model LF3 BXP. The reductions in RMSFE are mostly significant at short horizons in the mining and quarrying and raw material subcomponents. For the remaining PPI subcomponents, the results suggest that the low-frequency model does not produce significantly more accurate nowcasts than the mixed-frequency model. Indeed, for the clothing and daily sundry components, model LF3 BXP still generates forecasts with slightly smaller RMSFEs, although they are statistically insignificant. For the food component, LF3 BXP performs slightly worse than MF3 BXP from horizons 0 to 3. With regard to the durables component, LF3 BXP produces almost the same RMSFE at horizon 0 and slightly larger RMSFEs at horizons 1 to 6 than MF3 BXP.

Second, we compare the results of Mon BXP with those of LF3 BXP. For total PPI, Mon BXP produces larger RMSFEs than the baseline model (LF3 BXP) at all horizons. Similarly, the gains in nowcasting are mainly from the mining and quarrying, raw material, manufacturing, and clothing components, indicating the advantage of properly incorporating timely high-frequency data. For horizons 1 to 12, RMSFEs produced by Mon BXP are larger than RMSFEs produced by LF3 BXP for components such as mining and quarrying, raw materials, and manufacturing. In other components, such as food, clothing, daily sundries, and durables, Mon BXP produces smaller RMSFEs than LF3 BXP at medium to long horizons. This could be due to the lack of timely and higher frequency data in the consumer goods sectors. Moreover, we note that Mon BXP produces smaller RMSFEs than MF3 BXP at horizons 0 to 6 and larger RMSFEs than MF3 BXP at horizon 12. This indicates that the factor model using only monthly series can perform better than the mixed-frequency model in nowcasting and short-term forecasting.

Third, we compare the results of AR(1) and RW with those of the baseline model. For total PPI, AR(1) and RW are consistently and significantly outperformed by LF3 BXP at all horizons. For PPI subcomponents, AR(1) produces larger RMSFEs than LF3 BXP in most cases and for all horizons, except for the results of the food component and those of the daily sundry component at horizons 0, 1 and 6. RW in general has the worst performance among all models, especially for horizons other than 0. In a few exceptions, RW even outperforms MF3 BXP in nowcasting total PPI, mining and quarrying, and raw material components.

Although we choose the best model ex post (BXP) of LF3 as the baseline model, we also calculated the average RMSFEs produced by 24 factor models of LF3, which are characterized by different parameterizations (one to four factors and one to six lags) for reference. The results are shown in the last column of each block labeled LF3 Avg. We can see that LF3 Avg performs better than the alternative models at horizons 0 to 6 although slightly worse than AR(1) at horizons 6 and 12. Since we care more about short-term forecasts, a low-frequency dynamic factor model with transformations of high-frequency data can outperform other candidate models in nowcasting/forecasting total PPI inflation.

4.3 Subsample Forecasting Results

Next, we divide the full sample into two subsamples. The first subsample covers the period from January 2011 to December 2016, during which we use the same dataset as in Table 1, and the second subsample covers the period from January 2017 to December 2020, during which we add one additional type of higher frequency data, i.e., 10-day frequency data, to further compare the forecasting accuracy of the low-frequency and mixed-frequency factor models.

Table 2 reports the RMSFEs for the subsample from January 2011 to December 2016, and we can see that in this subsample, the results are quite similar to those of the full sample. In particular, LF3 BXP outperforms alternative models at almost all horizons for the total PPI and its mining and quarrying, raw material, manufacturing, and durables components.

Table 3 reports the RMSFEs for the subsample from January 2017 to December 2020, with two additional factor models using four types of data. In this table, the baseline model is LF4 BXP. We can see that in this subsample, LF4 BXP outperforms alternative models at all horizons for total PPI. We also calculate the average RMSFEs produced by the 24 factor models of LF4 and LF3, which are characterized by different parameterizations (one to four factors and one to six lags), and label them LF4 Avg and LF3 Avg. LF4 Avg performs better than LF3 Avg at all horizons, which demonstrates the high relevance of the 10-day data for total PPI. However, MF4 BXP performs worse than MF3 BXP for horizons from 0 to 3 and only outperforms at horizons 6 and 12. This seemingly contradictory result of utilizing high-frequency data to improve nowcasting is worth further investigation. We have two conjectures: one regarding the data patterns in particular and the other regarding the assumption implicitly imposed in the mixed-frequency dynamic factor model. High-frequency data that are relevant to China's PPIs are more volatile. Combining them into a mixed-frequency dynamic factor model without frequency transformations may amplify high-frequency fluctuations, thus producing more volatile low-frequency forecasts. The mixed-frequency factor model addresses lower frequency factors (i.e., weekly, monthly) observed at a daily frequency by implicitly treating them as stock variables, while a snapshot of PPI in the middle of the month can be a poor approximation of monthly average PPI inflation. For example, we calculated the month-on-month growth rate of the 11 daily RMP series, first with the snapshot price on the 15th of each month and second with the average price in each month and then treated the difference between the two as the nowcasting error. The RMSFEs, as a share of the mean absolute value of the month-on-month growth rates in each series, vary from 36% to 85%. The mean absolute value of the month-on-month growth rates of PPI from 2011:1 to 2020:12 is 0.39. This means that even the mixed-frequency model can exactly[8] predict the snapshot level of PPI; it will generate an RMSFE of 0.14 to 0.33, based on the assumption that the difference between the snapshot daily PPI index on the 15th of each month and the monthly average PPI index is approximately the same size as those of 11 RMP series. In fact, nowcasting the snapshot price using the mixed-frequency model will result in additional model prediction error; see the nowcasting RMSFE for the mixed-frequency BXP model reported in Table 1, which is 0.37, for reference.

5 Information Inflow and Forecasting Accuracy

In Section 4, we demonstrate the performance of various models, assuming that forecasts are produced once per month, i.e., on the 15th of each month. However, when conducting real-time forecasting, we encounter a continuous inflow of information, as new figures for various predictors are released nonsynchronously and with different degrees of delay. Therefore, we can produce a sequence of forecasts that are continuously updated when new data arrive. This section shows how information inflow in a data release is linked to the resulting forecast revision and how it can be used to evaluate the importance of different groups of variables to forecast, nowcast, and backcast total PPI inflation.

[8]Since the snapshot prices of RMP series are observed, ideally we can treat the true value as the forecasted value.

Figure 11 shows the evolution of RMSFEs computed every day, from 48 days before the beginning of the reference month to 45 days after the beginning of the reference month. The y-axis measures the RMSFEs of PPI forecasts calculated over the full sample from January 2011 to December 2020 for various models.[9] The x-axis is the 93 days of the "prediction period" for each month. The models' monthly PPI inflation prediction is first made 48 days before the start of the reference month. For illustration, suppose that the reference month is July and that the first forecast is made when the April PPI data are released. Then, it is updated with each successive data release until the release of the PPI of the reference month, which takes place 45 days after the start of the current month. Thus, the prediction period for each PPI release covers four months (two previous months for forecasting, one current month for nowcasting, and one coming month for backcasting).

We can see from Figure 11 that the RMSFEs produced by the different models all decline gradually over the prediction period, which means that new information can indeed help to improve prediction accuracy. We compare the results of the low-frequency factor model and the mixed-frequency factor model first. On each day of the prediction period, LF3 BXP always performs better than MF3 BXP, which is consistent with our conclusion in Section 4. Then, we consider other alternative models. AR performs better than RW throughout the timeline. At the early stage of the prediction period, Mon BXP exhibits almost the same accuracy as MF3 BXP but performs worse than LF3 BXP. On day -31, when the PMI data two months before the reference month were released, Mon BXP performed better than MF3 BXP and almost the same as LF3 BXP. However, since LF3 BXP can update forecasts using higher frequency information, it outperforms Mon BXP from day -31 to day 0 and keeps performing relatively better than Mon BXP from day 0 to day 30. On day 31, when the PMI data of the current month are released, Mon BXP exceeds LF3 BXP consistently until the end of the prediction period. Overall, LF3 BXP outperforms other alternative models in the forecasting and nowcasting periods and only performs slightly worse than Mon BXP in the backcasting period.

Similarly, Figure 12 displays the evolution of forecasting accuracy over the subsample from January 2017 to December 2020 for an additional type of data. Robustly speaking, the RMSFEs of LF3 BXP and LF4 BXP are smaller than those of MF3 BXP and MF4 BXP on almost all days in the prediction period, except for the first day, i.e., on day -47, the RMSFE

[9] Taking the RMSFE at t=1 as an example, forecasts of PPI in each month are calculated on the first day of the reference month. Then, the RMSFE at that point is the deviation of those forecasts from the true values.

of MF3 BXP is approximately the same as the RMSFE of LF3 BXP. Before day 15, LF4 BXP performs better than LF3 BXP. However, after day 15, LF3 BXP performed slightly better than LF4 BXP. This is also confirmed in Table 3, where the RMSFE of LF3 BXP at horizon 0 is 0.23, just slightly larger than the RMSFE of LF4 BXP at horizon 0, which is 0.22. Model Mon BXP can always outperform models AR(1) and RW and outperforms the mixed-frequency models after day 0 but always underperforms the low-frequency model LF4 BXP before day 31 and LF3 BXP. Overall, LF4 BXP outperforms other alternative models on almost all days in the prediction period and is slightly worse than Mon BXP and LF3 BXP only for a short period. Low-frequency DFM exhibits consistent superior forecasting/nowcasting power over different samples and improves as more high-frequency data become available.

6 Conclusion

This paper compares the forecasting performance of a low-frequency dynamic factor model and a mixed-frequency dynamic factor model, which both exploit high-frequency information to forecast China's total PPI inflation and its components. The paper focuses on five groups of data, including sampling frequencies that are higher than monthly. The first group includes 11 daily market prices of raw materials. The second consists of 10 weekly commodity price indexes. The third group contains 47 series of producer goods market prices released at a 10-day frequency. The fourth group is the monthly PMI Purchasing Price Index, which is released earlier and comoves closely with PPI inflation. The fifth group collects some daily financial variables that are believed to be relevant. Forecasts of the target variable are calculated in dynamic factor model frameworks in which mixed-frequency data are utilized in different ways. Similar to Lenza and Warmedinger (2011), the low-frequency factor model estimates monthly factors by taking averages of the high-frequency data to obtain monthly indicators. As in Modugno (2013), the mixed-frequency factor model enables a unified framework for data at all frequencies and allows restrictions on the coefficients, such as the time-varying coefficients, to aggregate the daily factors. Our empirical results for China's PPI inflation suggest that the low-frequency factor model generally performs better than the mixed-frequency approach for total PPI. This superiority is especially significant for nowcasting at horizon 0 and short-term forecasting. The gains in forecasting accuracy are mainly due to improvements from the mining and quarrying, raw material, and manufacturing components. The results are quite robust to the consideration of different subsamples and the inclusion additional types of data. Moreover, we evaluate the performance of all our models through the timeliness of data releases. The evidence shows that all models improve gradually as new information arrives. Models utilizing high-frequency data can provide continuous revisions to real-time forecasting. Overall, the simple and parsimonious low-frequency factor model has the best performance throughout the timeline in forecasting China's total PPI inflation.

References

Andreou, Elena, Eric Ghysels, and Andros Kourtellos, "Should macroeconomic forecasters use daily financial data and how?," *Journal of Business & Economic Statistics*, 2013, *31* (2), 240–251.

Bańbura, Marta and Michele Modugno, "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data," *Journal of Applied Econometrics*, 2014, *29* (1), 133–160.

Diebold, Francis X. and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 1995, *13* (3), 253–263.

Gali, Jordi and Tommaso Monacelli, "Monetary policy and exchange rate volatility in a small open economy," *The Review of Economic Studies*, 2005, *72* (3), 707–734.

Giannone, Domenico, Lucrezia Reichlin, and David Small, "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 2008, *55* (4), 665–676.

Huang, Kevin XD and Zheng Liu, "Inflation targeting: What inflation rate to target?," *Journal of Monetary Economics*, 2005, *52* (8), 1435–1462.

Knotek, Edward S and Saeed Zaman, "Nowcasting US headline and core inflation," *Journal of Money, Credit and Banking*, 2017, *49* (5), 931–968.

Lenza, Michele and Thomas Warmedinger, "A factor model for euro-area short-term inflation analysis," *Jahrbücher für Nationalökonomie und Statistik*, 2011, *231* (1), 50–62.

Modugno, Michele, "Now-casting inflation using high frequency data," *International Journal of Forecasting*, 2013, *29* (4), 664–675.

Monteforte, Libero and Gianluca Moretti, "Real-time forecasts of inflation: The role of financial variables," *Journal of Forecasting*, 2013, *32* (1), 51–61.

Paoli, Bianca De, "Monetary policy and welfare in a small open economy," *Journal of International Economics*, 2009, *77* (1), 11–22.

Wei, Shang-Jin and Yinxi Xie, "The Wedge of the Century: Understanding a Divergence between CPI and PPI Inflation Measures," Technical Report, National Bureau of Economic Research 2018.

_ **and** _ , "Monetary policy in an era of global supply chains," *Journal of International Economics*, 2020, *124*, 103299.

Table 1
Forecast evaluation: full sample (January 2011–December 2020).

Horizon	LF3 BXP	MF3 BXP	Mon BXP	AR(1)	RW	LF3 Avg	LF3 BXP	MF3 BXP	Mon BXP	AR(1)	RW	LF3 Avg
	Total						Mining and Quarrying					
12	3.37	**3.88***	**4.24*****	**3.61**	**6.06*****	**4.00**	10.82	**12.34**	**13.78***	**12.10**	**22.20*****	**12.61**
6	2.15	**2.70**	**2.52****	**2.44***	**3.61*****	**2.45**	7.80	**9.68***	**8.85**	**8.55**	**13.86*****	**8.14**
3	1.23	**1.52**	**1.44****	**1.52*****	**1.95*****	**1.40**	4.89	**6.14***	**5.68***	**5.88**	**8.21*****	**5.03**
1	0.54	**0.70****	**0.69*****	**0.74*****	**0.84*****	**0.62**	2.44	**3.45*****	**2.91****	**3.20****	**3.81*****	**2.54**
0	0.25	**0.37*****	**0.28***	**0.32*****	**0.34*****	**0.28**	1.40	**1.95*****	1.33	**1.52**	**1.66**	**1.41**
	Raw Material						Manufacturing					
12	6.30	**7.08****	**7.64****	**6.77**	**12.08*****	**7.10**	3.20	**3.59**	**3.72****	**3.31**	**5.29*****	**3.62**
6	4.16	**4.97***	**4.74****	**4.64***	**7.60*****	**4.45**	1.93	**2.29**	**2.18***	**2.02**	**2.93*****	**2.16**
3	2.52	**2.94***	**2.93****	**3.09*****	**4.25*****	**2.68**	1.08	**1.24**	**1.21**	**1.22****	**1.57*****	**1.21**
1	1.23	**1.46****	**1.50*****	**1.62*****	**1.89*****	**1.31**	0.49	**0.53**	**0.60*****	**0.62*****	**0.72*****	**0.54**
0	0.61	**0.79*****	**0.67**	**0.73*****	**0.76*****	**0.63**	0.23	**0.27**	**0.28*****	**0.30*****	**0.31*****	**0.25**
	Food						Clothing					
12	2.86	**2.91**	2.74	2.67	**4.08**	**3.04**	1.33	**1.36**	1.29	**1.71**	**2.05*****	**1.43**
6	1.88	**1.97**	1.67**	1.57	**2.26**	**1.94**	0.92	**0.98**	0.86	**1.09**	**1.20***	**0.95**
3	1.17	1.15	0.99***	0.91***	**1.27**	1.17	0.59	**0.62**	0.57	**0.66**	**0.64**	**0.60**
1	0.62	0.59	0.54***	0.46***	0.60	0.60	0.33	**0.35**	**0.35**	**0.36**	**0.33**	**0.34**
0	0.34	0.32	0.29***	0.24***	0.26**	0.32	0.20	**0.20**	**0.26*****	**0.22**	0.17	0.19
	Daily Sundries						Durables					
12	1.46	**1.69**	1.43	**1.46**	**2.33*****	**1.63**	0.64	**0.68**	0.60*	**1.06*****	**1.87*****	**0.65**
6	0.89	**1.03**	0.79**	0.88	**1.29*****	**0.96**	0.47	0.46	0.45	**0.69*****	**1.00*****	**0.47**
3	0.53	**0.58**	0.44**	**0.55**	**0.74****	**0.56**	0.34	0.32	0.33	**0.44*****	**0.60*****	0.33
1	0.30	**0.32**	0.27**	0.30	**0.36**	**0.31**	0.23	0.22	**0.24**	**0.26*****	**0.33*****	0.22
0	0.18	**0.18**	0.17	0.16	**0.18**	0.18	0.15	0.15*	**0.16**	**0.16**	**0.18****	0.15

Notes: LF3 BXP denotes the best model ex post of the low-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 4 factors and 1 lag; MF3 BXP denotes the best model ex post of the mixed-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 3 factors and 31 lags; Mon BXP denotes the best model ex post of the dynamic factor model with only monthly frequency series, i.e., the factor model with 3 factors and 2 lags; AR(1) and RW denote the autoregressive model with lag order 1 and the random walk model, respectively; and LF3 Avg denotes the average of the RMSFEs produced by the low-frequency factor models with 3 types of data frequencies, with 1 to 4 factors in Eq. (1), and from 1 to 6 lags in Eq.(2).
*, **, and *** denote rejection of the null hypothesis of the equal predictability test in Diebold and Mariano (1995) of the baseline model (LF3 BXP) versus each alternative model at the 10%, 5%, and 1% level, respectively. **Bold** entries denote that the alternative model produces larger RMSFEs than the baseline model.
We do not perform the DM test on LF3 Avg, since it is an average of 24 models.

Table 2

Forecast evaluation: subsample (January 2011–December 2016).

Horizon	LF3 BXP	MF3 BXP	Mon BXP	AR(1)	RW	LF3 Avg	LF3 BXP	MF3 BXP	Mon BXP	AR(1)	RW	LF3 Avg
	Total						Mining and Quarrying					
12	3.19	**3.84**	**4.56*** **	**3.60**	**5.69** **	**4.30**	9.95	**11.40**	**14.88*** **	**11.99* **	**18.15*** **	**13.30**
6	2.05	**2.81**	**2.55** **	**2.25**	**3.08*** **	**2.58**	7.71	**10.27* **	**9.20**	**8.33**	**11.88** **	**8.41**
3	1.19	**1.62**	**1.34**	**1.36* **	**1.66*** **	**1.45**	4.70	**6.81** **	**5.69**	**5.53**	**6.97* **	**5.05**
1	0.54	**0.76** **	**0.66* **	**0.67** **	**0.73*** **	**0.64**	2.31	**3.93*** **	**2.81**	**2.83**	**3.16 ** **	**2.52**
0	0.27	**0.38*** **	0.26	**0.29**	**0.30**	**0.30**	1.48	**2.10*** **	1.29	1.31	1.39	1.48
	Raw Material						Manufacturing					
12	5.38	**6.40* **	**7.73*** **	**6.42**	**10.79*** **	**6.96**	2.95	**3.53**	**3.75*** **	**3.28**	**4.96* **	**3.70**
6	3.64	**4.82**	**4.50*** **	**4.20**	**6.19*** **	**4.27**	1.85	**2.35**	**2.12* **	**1.94**	**2.49** **	**2.24**
3	2.27	**2.92* **	**2.66** **	**2.73** **	**3.50*** **	**2.56**	1.05	**1.26**	**1.09**	**1.15**	**1.35** **	**1.24**
1	1.11	**1.47*** **	**1.36** **	**1.40*** **	**1.55*** **	**1.22**	0.51	**0.54**	**0.59**	**0.63** **	**0.68*** **	**0.56**
0	0.58	**0.76*** **	**0.59**	**0.65**	**0.64**	**0.60**	0.24	**0.26**	**0.27**	**0.29* **	**0.28**	**0.26**
	Food						Clothing					
12	3.31	**3.37**	3.08***	2.94	**4.04**	**3.54**	1.18	**1.28**	**1.24**	**2.01**	**1.96*** **	**1.35**
6	2.07	**2.19**	1.65***	1.58	1.99	**2.16**	0.81	**0.98**	0.73	**1.23* **	**1.17*** **	**0.85**
3	1.21	1.19	0.78***	0.81***	1.07	**1.22**	0.50	**0.62* **	0.46	**0.72* **	**0.63* **	**0.52**
1	0.61	0.56	0.38***	0.35***	0.48**	0.58	0.29	**0.36** **	**0.30**	**0.38** **	**0.35**	**0.30**
0	0.33	0.30	0.20***	0.18***	0.20***	0.30	0.18	**0.20**	**0.24*** **	**0.23** **	0.18	0.17
	Daily Sundries						Durables					
12	1.76	**2.08**	1.74	1.75	**2.39**	**1.99**	0.43	**0.53**	0.39	**0.92*** **	**1.90*** **	**0.43**
6	1.04	**1.26**	0.89***	1.02	**1.17**	**1.14**	0.37	**0.38**	**0.39**	**0.60*** **	**0.99*** **	0.36
3	0.60	**0.69**	0.44***	0.58	**0.61**	**0.63**	0.29	**0.29**	**0.32**	**0.38*** **	**0.58*** **	0.28
1	0.33	**0.35**	0.27**	0.31	**0.35**	**0.34**	0.20	**0.20**	**0.23**	**0.23** **	**0.30*** **	0.19
0	0.19	0.18	0.16	0.16	0.16	0.18	0.14	0.13	**0.16**	**0.15**	**0.17** **	0.13

Notes: LF3 BXP denotes the best model ex post of the low-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 4 factors and 1 lag; MF3 BXP denotes the best model ex post of the mixed-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 3 factors and 31 lags; Mon BXP denotes the best model ex post of the dynamic factor model with only monthly frequency series, i.e., the factor model with 3 factors and 2 lags; AR(1) and RW denote the autoregressive model with lag order 1 and the random walk model, respectively; and LF3 Avg denotes the averages of the RMSFEs produced by the low-frequency factor models with 3 types of data frequencies, with 1 to 4 factors in Eq.(1), and from 1 to 6 lags in Eq.(2).
*, **, and *** denote rejection of the null hypothesis of the equal predictability test in Diebold and Mariano (1995) of the baseline model (LF3 BXP) versus each alternative model at the 10%, 5%, and 1% level, respectively. **Bold** entries denote that the alternative model produces larger RMSFEs than the baseline model.
We do not perform the DM test on LF3 Avg, since it is an average of 24 models.

Table 3

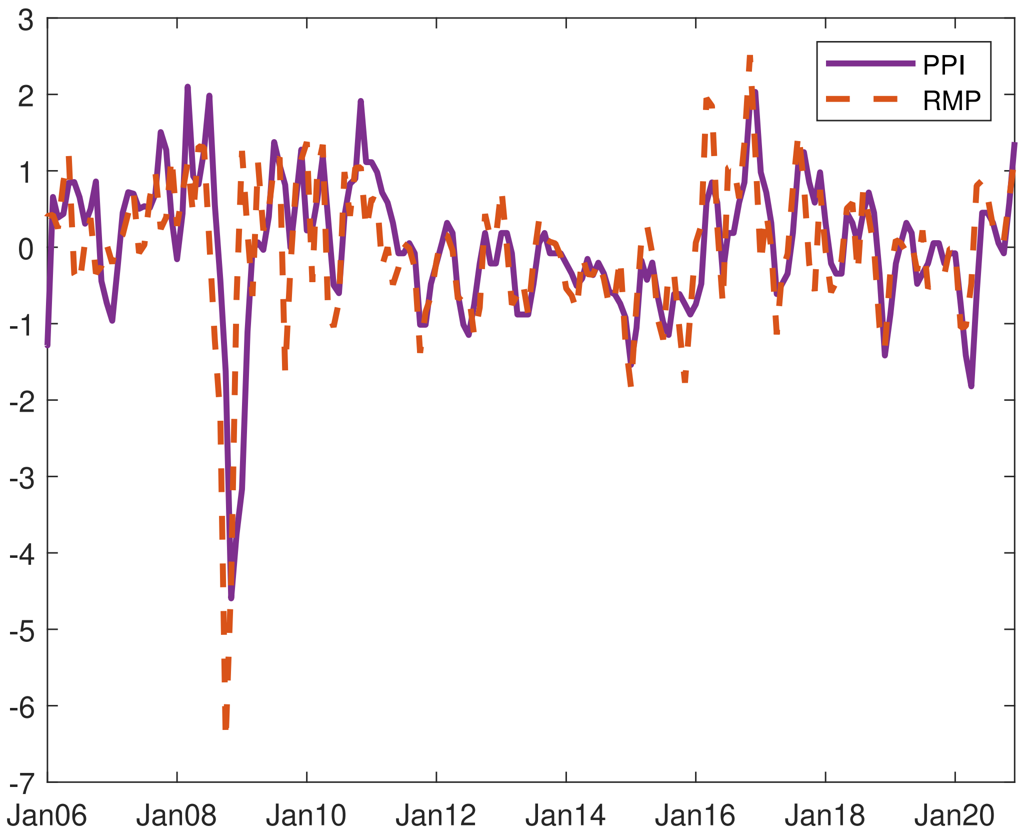
Forecast evaluation: subsample (January 2017–December 2020).

Horizon	LF4 BXP	LF3 BXP	MF4 BXP	MF3 BXP	Mon BXP	AR(1)	RW	LF4 Avg	LF3 Avg
	Total								
12	3.02	**3.61***	**3.76***	**3.94*****	**3.72**	**3.63**	**6.57*****	**3.16**	**3.52**
6	1.73	**2.28****	**2.26**	**2.52****	**2.47**	**2.69****	**4.27*****	**1.99**	**2.23**
3	1.09	**1.30****	**1.38**	**1.37**	**1.57****	**1.72*****	**2.32*****	**1.16**	**1.31**
1	0.48	**0.56**	**0.68*****	**0.58**	**0.73*****	**0.83*****	**0.99*****	**0.53**	**0.58**
0	0.22	**0.23**	**0.39*****	**0.36****	**0.31*****	**0.37*****	**0.40*****	**0.25**	**0.25**
	Raw Material								
12	6.51	**7.48***	**7.64***	**7.99*****	**7.50**	**7.26**	**13.78*****	**6.70**	**7.29**
6	3.88	**4.83*****	**4.85***	**5.19*****	**5.07**	**5.24****	**9.32*****	**4.31**	**4.70**
3	2.53	**2.86****	**3.15****	**2.98**	**3.30****	**3.56*****	**5.18*****	**2.68**	**2.85**
1	1.28	**1.40**	**1.71*****	**1.45**	**1.68*****	**1.90*****	**2.32*****	**1.40**	**1.43**
0	0.62	**0.65**	**0.91*****	**0.84****	**0.77****	**0.85*****	**0.92****	**0.71**	**0.68**
	Food								
12	2.18	2.01**	1.92*	2.03	2.15	**2.19**	**4.14**	2.15	2.06
6	1.67	1.55**	1.42*	1.57	**1.71**	1.54	**2.61**	1.61	1.55
3	1.12	1.11	1.00**	1.11	**1.25***	1.05	**1.52**	1.10	1.09
1	0.63	**0.65**	0.60*	**0.63**	**0.72****	0.60	**0.73**	0.63	0.63
0	0.35	**0.35**	0.35	**0.36**	**0.39****	0.31	0.34	**0.35**	0.35
	Daily Sundries								
12	0.85	0.83	0.84	0.82	0.78	**0.86**	**2.24*****	0.84	0.84
6	0.63	0.60	0.58	0.55	0.61	0.60	**1.45*****	0.59	0.61
3	0.41	0.40	**0.44**	0.38	**0.44**	**0.50**	**0.91*****	0.40	**0.42**
1	0.25	**0.25**	**0.27***	**0.25**	**0.26**	**0.27**	**0.37****	**0.25**	**0.26**
0	0.17	**0.17**	**0.18****	**0.18**	**0.17**	**0.17**	**0.21**	0.16	**0.17**

Horizon	LF4 BXP	LF3 BXP	MF4 BXP	MF3 BXP	Mon BXP	AR(1)	RW	LF4 Avg	LF3 Avg
	Mining and Quarrying								
12	10.03	**12.00**	**11.66**	**13.62*****	**11.96**	**12.26**	**27.16*****	9.81	**11.49**
6	6.83	**7.93**	**7.08**	**8.73**	**8.30**	**8.85**	**16.40****	**6.98**	**7.71**
3	4.88	**5.15**	4.66	**4.96**	**5.68**	**6.37**	**9.78****	4.74	**5.00**
1	2.47	**2.61**	**2.78**	**2.56**	**3.05***	**3.68****	**4.62****	**2.58**	**2.57**
0	1.25	**1.27**	**1.75***	**1.71***	**1.39**	**1.79****	**1.99****	**1.37**	**1.30**
	Manufacturing								
12	3.10	**3.54***	**3.74**	**3.69*****	**3.69**	**3.35**	**5.75*****	**3.25**	**3.49**
6	1.57	**2.05****	**2.13***	**2.20*****	**2.26***	**2.13**	**3.50*****	**1.83**	**2.03**
3	0.92	**1.11****	**1.25***	**1.19****	**1.37****	**1.32****	**1.85*****	**1.02**	**1.15**
1	0.42	**0.46**	**0.57*****	**0.50**	**0.62*****	**0.60*****	**0.77*****	**0.45**	**0.51**
0	0.21	**0.21**	**0.30*****	**0.29***	**0.30*****	**0.30*****	**0.35*****	**0.21**	**0.24**
	Clothing								
12	1.54	**1.54**	1.46	1.46	1.36**	1.12	**2.18**	1.51	1.53
6	1.10	1.08	1.03	0.97***	1.02	0.83	**1.26**	1.05	1.07
3	0.71	0.70	0.67	0.61***	0.69	0.56	0.66	0.69	0.71
1	0.38	0.38	0.38	0.34***	**0.41**	0.31	0.31**	0.38	**0.39**
0	0.22	**0.22**	**0.22**	0.19***	**0.29*****	0.19	0.17*	0.21	**0.22**
	Durables								
12	0.91	0.87***	0.87**	0.85	0.81***	**1.24**	**1.83*****	**0.91**	0.88
6	0.61	0.59***	0.58**	0.55**	0.53***	**0.80**	**1.03*****	0.61	0.60
3	0.40	0.40	0.38**	0.36**	0.35**	**0.51****	**0.62*****	**0.41**	0.40
1	0.26	**0.27**	0.25**	0.24***	0.26	**0.31****	**0.37*****	**0.26**	**0.26**
0	0.17	**0.17***	0.16	0.16	**0.17**	**0.19****	**0.20**	**0.17**	**0.17**

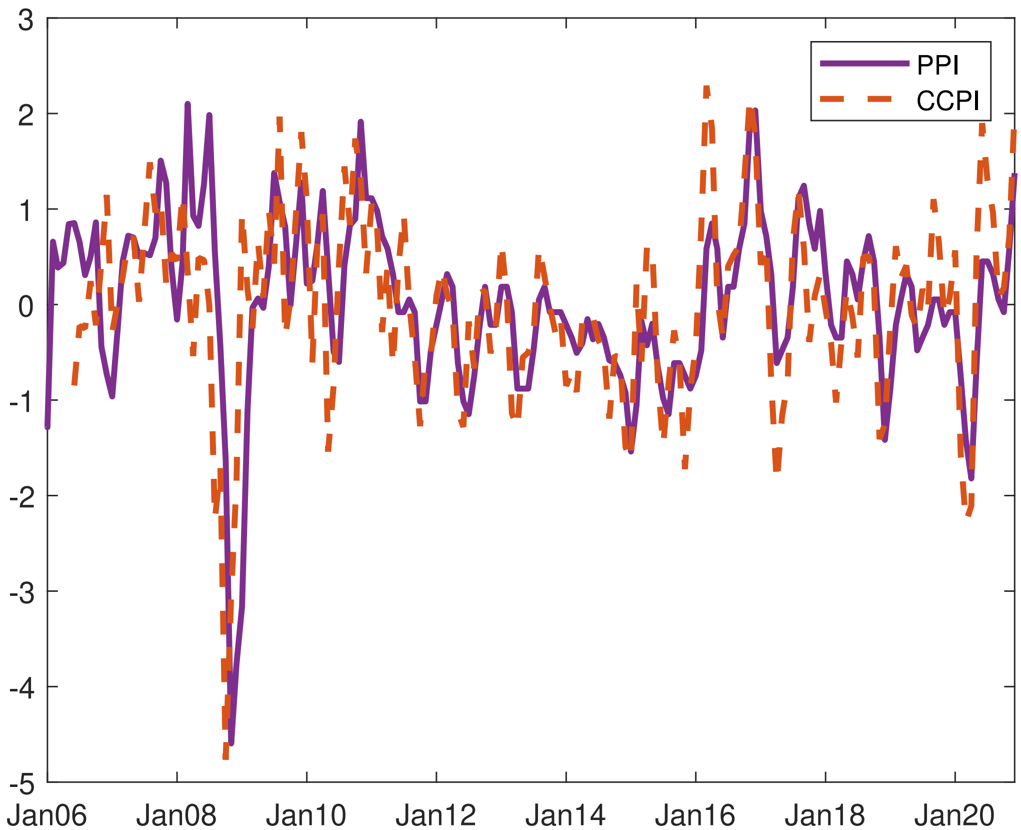
Notes: LF4 BXP denotes the best model ex post of the low-frequency dynamic factor model with 4 types of data frequencies, i.e., the factor model with 2 factors and 6 lags; LF3 BXP denotes the best model ex post of the low-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 4 factors and 1 lag; MF4 BXP denotes the best model ex post of the mixed-frequency dynamic factor model with 4 types of data frequencies, i.e., the factor model with 1 factor and 3 lags; MF3 BXP denotes the best model ex post of the mixed-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 3 factors and 31 lags; Mon BXP denotes the best model ex post of the dynamic factor model with only monthly frequency series, i.e., the factor model with 3 factors and 2 lags; and AR(1) and RW denote the autoregressive model with lag order 1 and the random walk model, respectively. LF4 Avg denotes the averages of the RMSFEs produced by the low-frequency factor models with 4 types of data frequencies, with 1 to 4 factors in Eq. (1), and from 1 to 6 lags in Eq. (2). LF3 Avg denotes the averages of the RMSFEs produced by the low-frequency factor models with 3 types of data frequencies, with 1 to 4 factors in Eq. (1), and from 1 to 6 lags in Eq. (2). *, **, and *** denote rejection of the null hypothesis of the equal predictability test in Diebold and Mariano (1995) of the baseline model (LF4 BXP) versus each alternative model at the 10%, 5%, and 1% level, respectively. **Bold** entries denote that the alternative model produces larger RMSFEs than the baseline model. We do not perform the DM test on LF3 Avg and LF4 Avg, since each is the average of 24 models.

Figure 1: Raw material prices



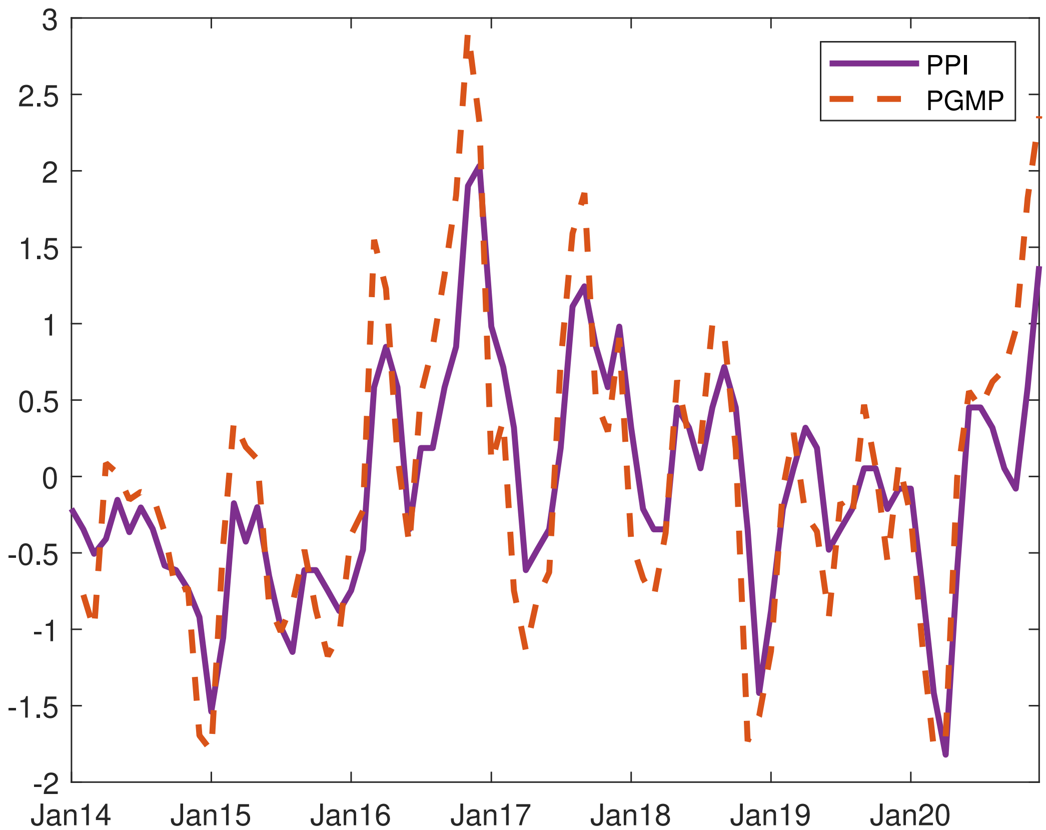
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the average month-on-month growth rates of eleven raw material price (RMP) series. The original RMP series are in daily frequency and transformed to monthly frequency by taking the monthly average. All of the series are standardized and centered on zero.

Figure 2: China commodity price index



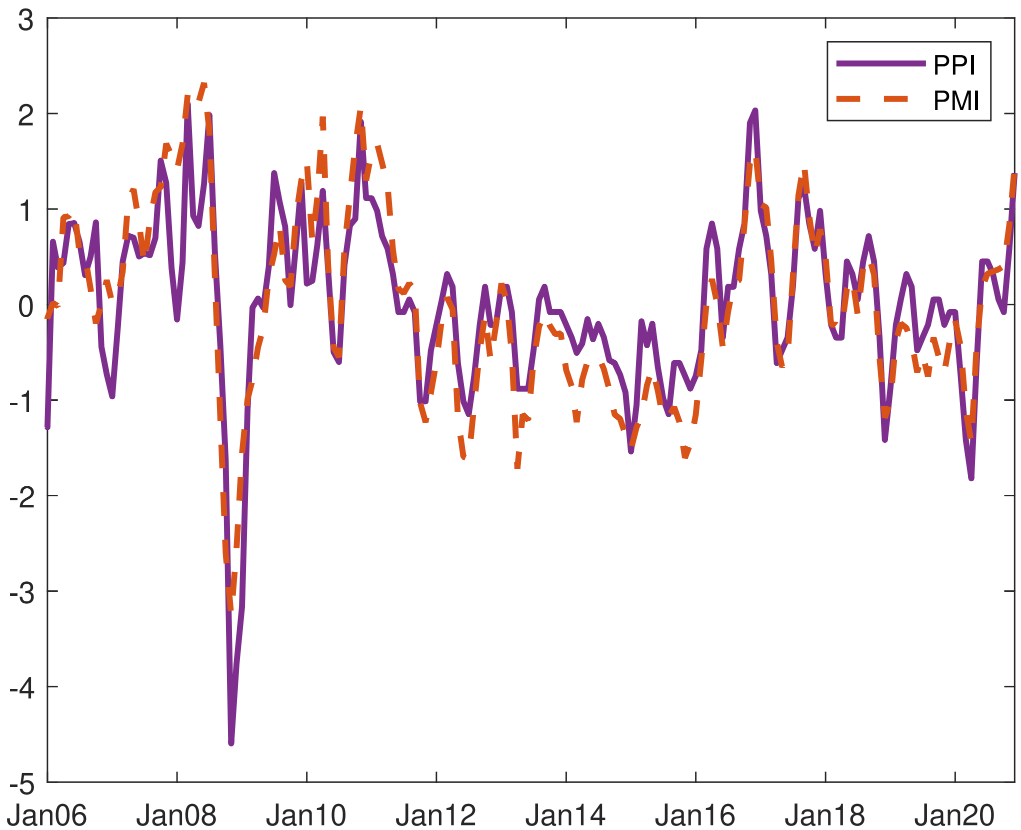
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the average month-on-month growth rates of eight China Commodity Price Index (CCPI) series. The original CCPI series are in weekly frequency and transformed to a monthly frequency by taking the monthly average. CCPI data began to be released in June 2006. All of the series are standardized and centered on zero.

Figure 3: Producer goods market prices



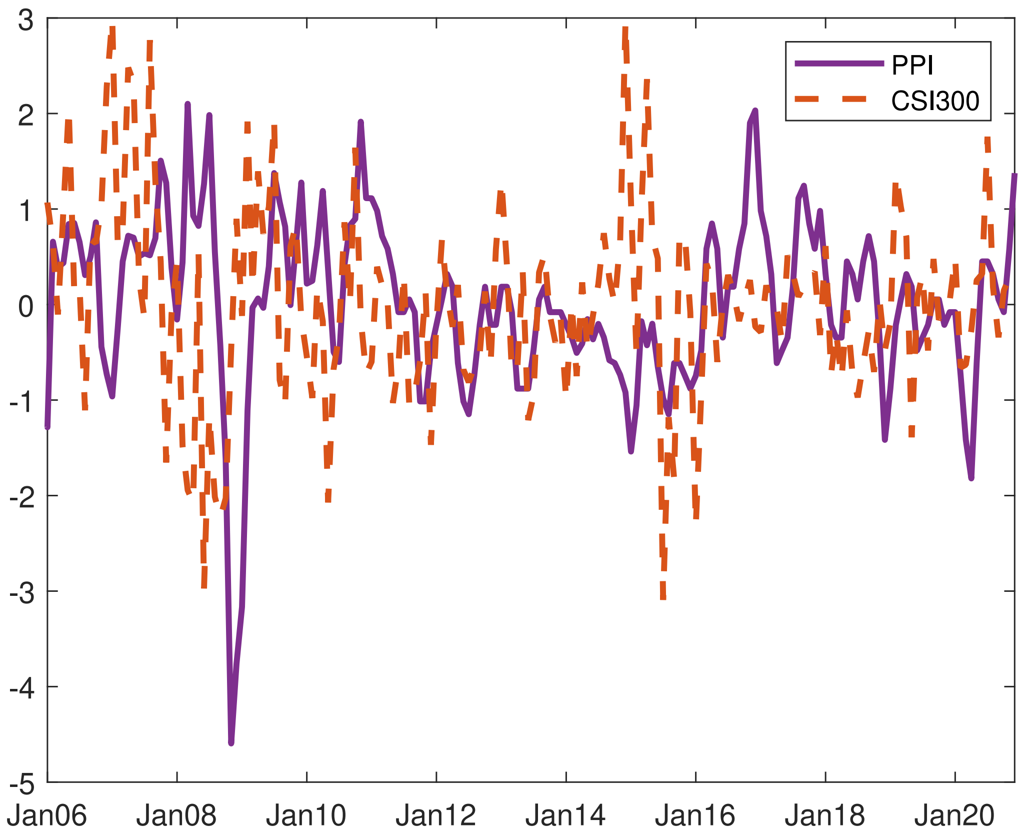
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the average month-on-month growth rates of forty-seven producer goods market price (PGMP) series. The original CCPI series has a 10-day frequency and is transformed to a monthly frequency by taking the monthly average. Note that PGMP data began to be released in January 2014, so the month-on-month growth rates of PGMP start in February 2014. All of the series are standardized and centered on zero.

Figure 4: PMI



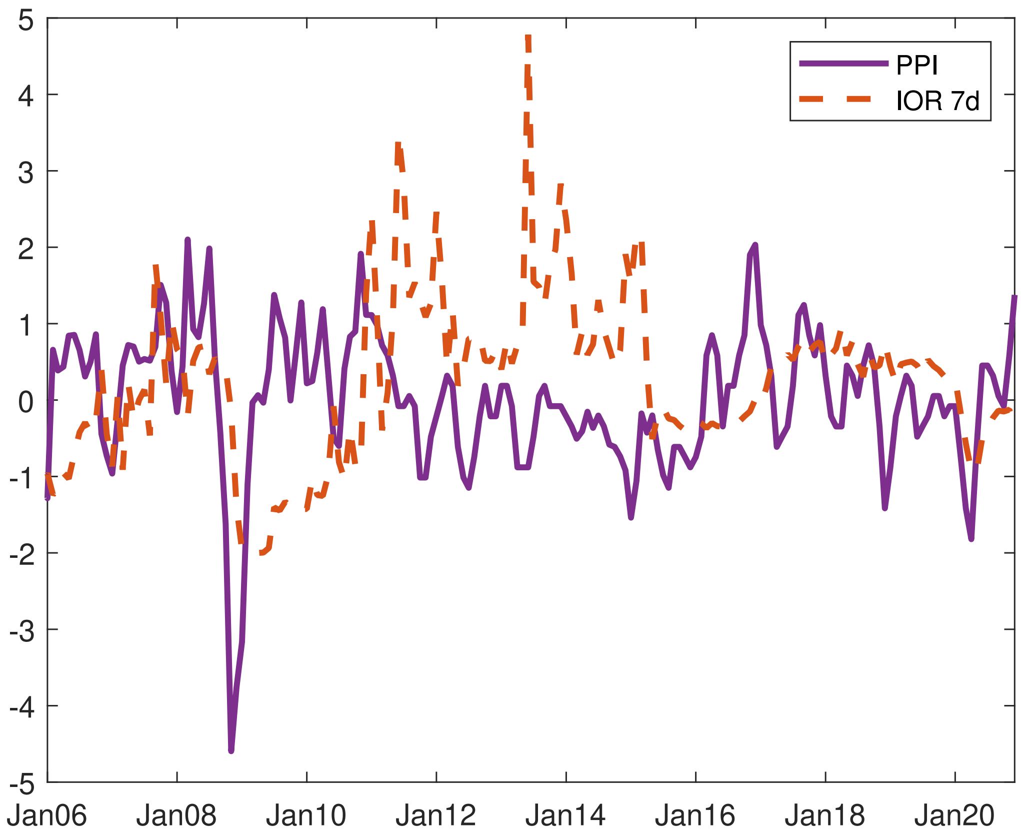
Notes: This figure shows the evolution of the month-on-month growth rates of the overall PPI and PMI subindex: Purchasing Price Index. All of the series are standardized and centered on zero.

Figure 5: CSI 300 stock index



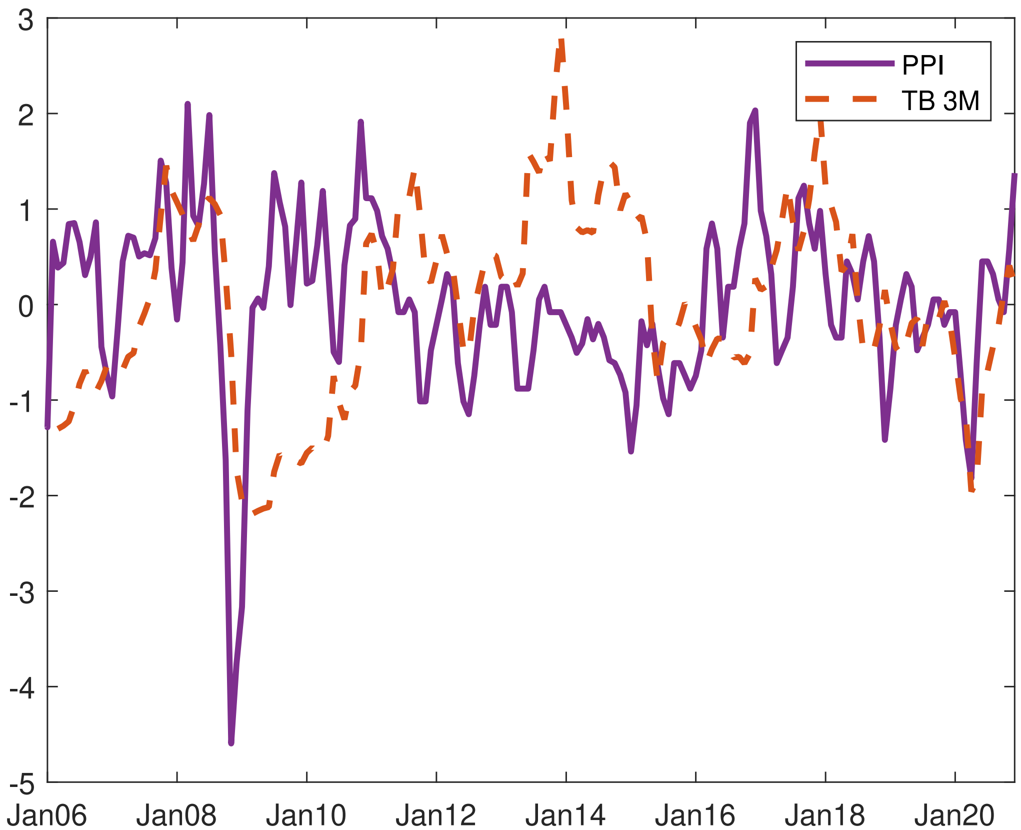
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the month-on-month growth rates of the CSI 300 stock index (CSI 300). The original CSI 300 stock index is at a daily frequency and transformed to monthly frequency by taking the monthly average. All of the series are standardized and centered on zero.

Figure 6: 7-day weighted average interbank offered rate



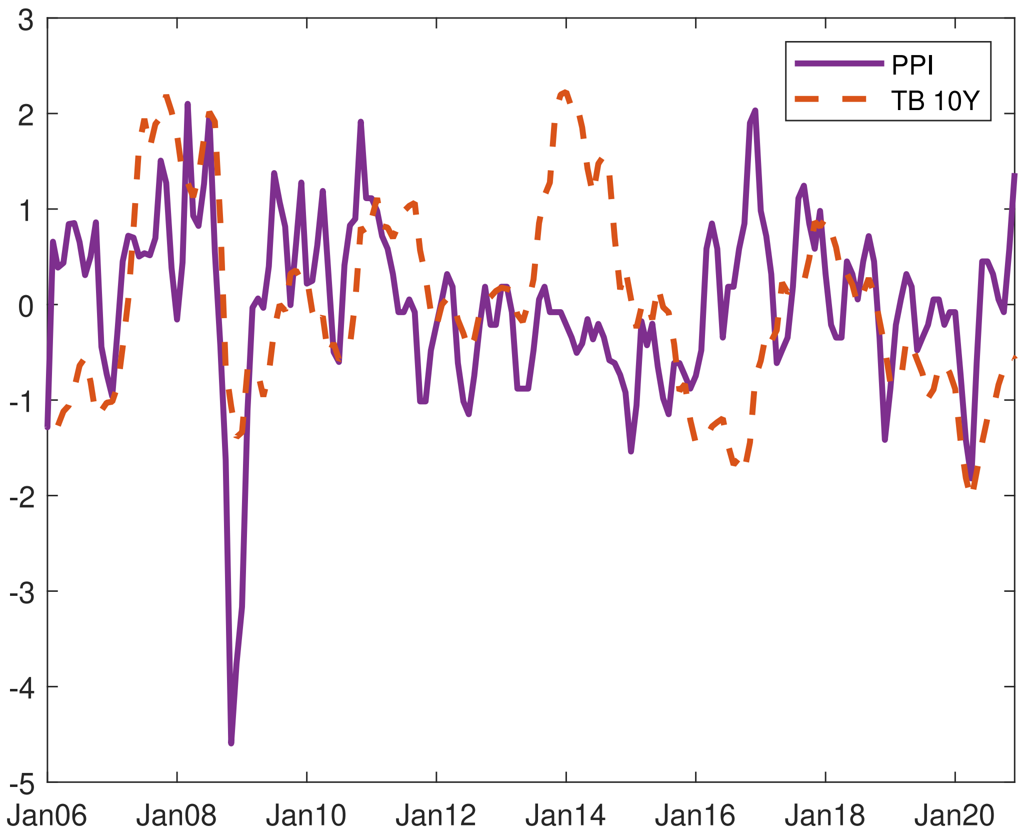
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the 7-day weighted average interbank offered rate (IOR7d). The original IOR7d is at a daily frequency and transformed to monthly frequency by taking the monthly average. All of the series are standardized and centered on zero.

Figure 7: Three-month Treasury bond rate



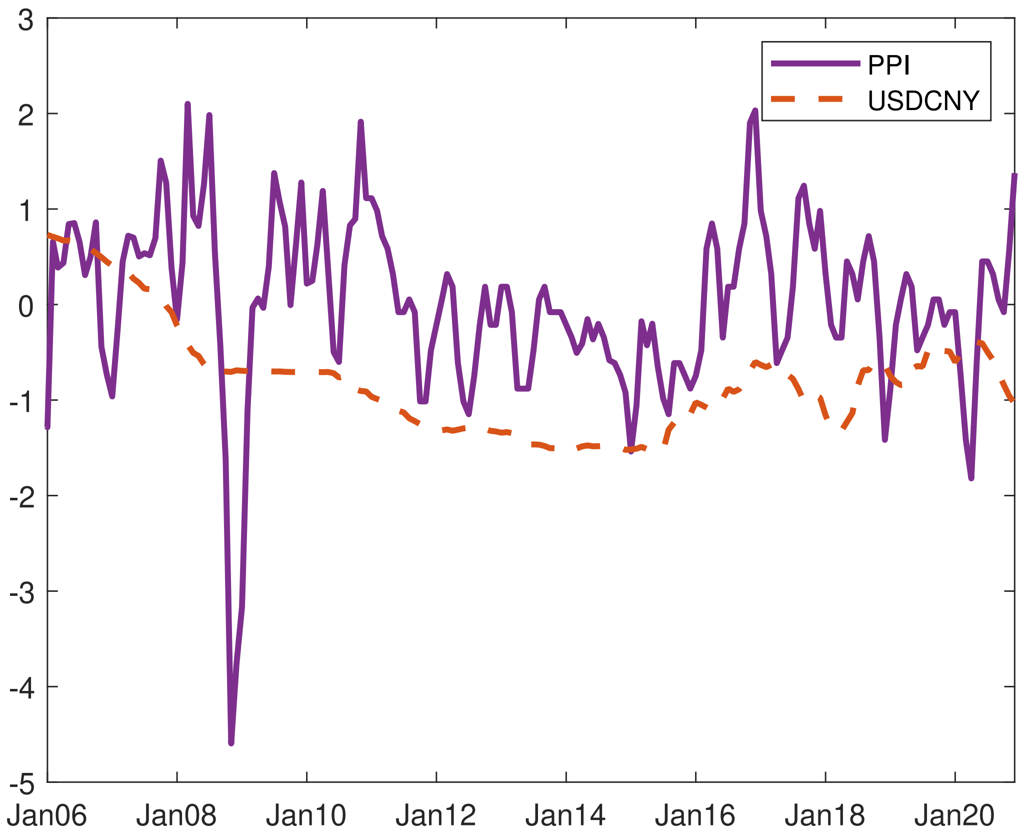
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the three-month Treasury bond rate (TB3M). The original TB3M is at a daily frequency and transformed to monthly frequency by taking the monthly average. Note that TB3M began to be released in March 2006. All of the series are standardized and centered on zero.

Figure 8: 10-year Treasury bond rate



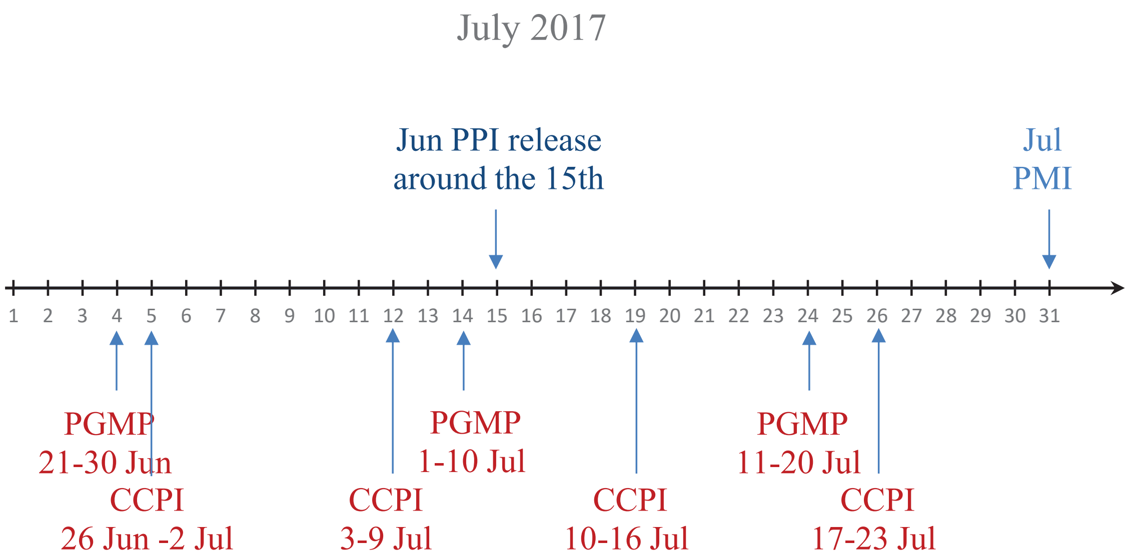
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the ten-year Treasury bond rate (TB10Y). The original TB10Y is at a daily frequency and transformed to monthly frequency by taking the monthly average. Note that TB10Y began to be released in March 2006. All of the series are standardized and centered on zero.

Figure 9: CNY central parity rate of USD



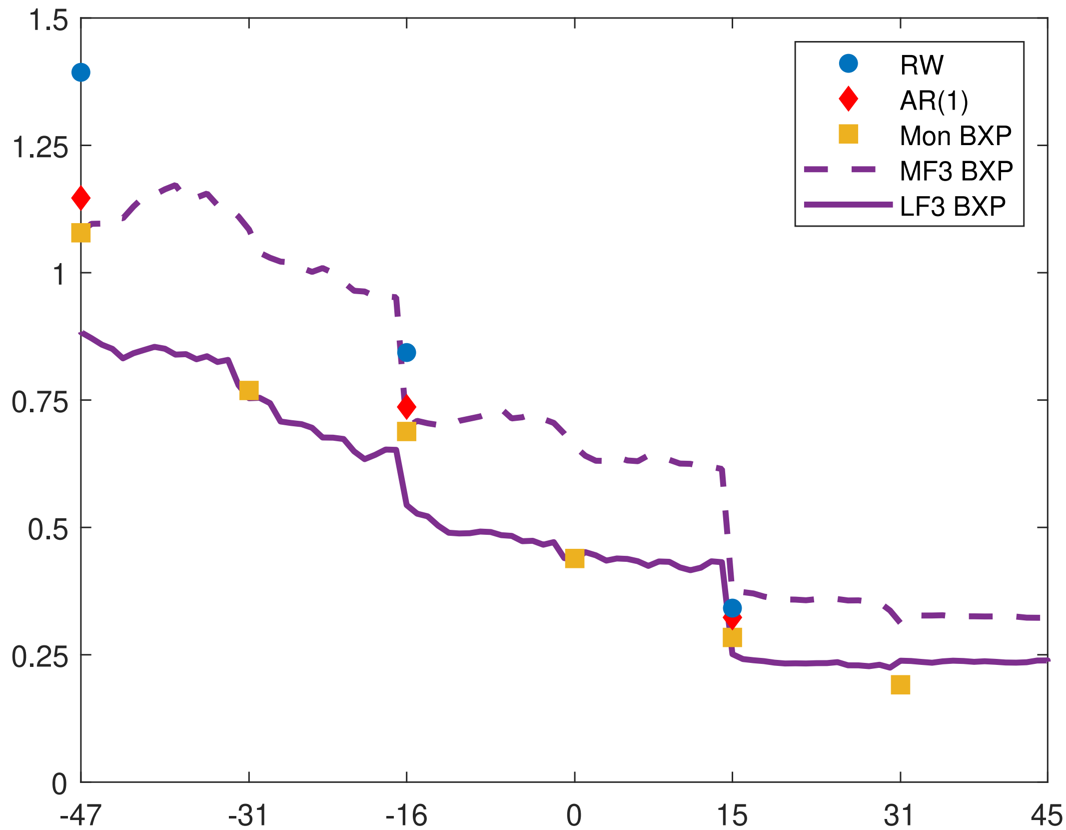
Notes: This figure shows the evolution of the month-on-month growth rates of overall PPI and the central parity rate of RMB against USD (USDCNY). All of the series are standardized and centered on zero.

Figure 10: Timeliness of Data Release



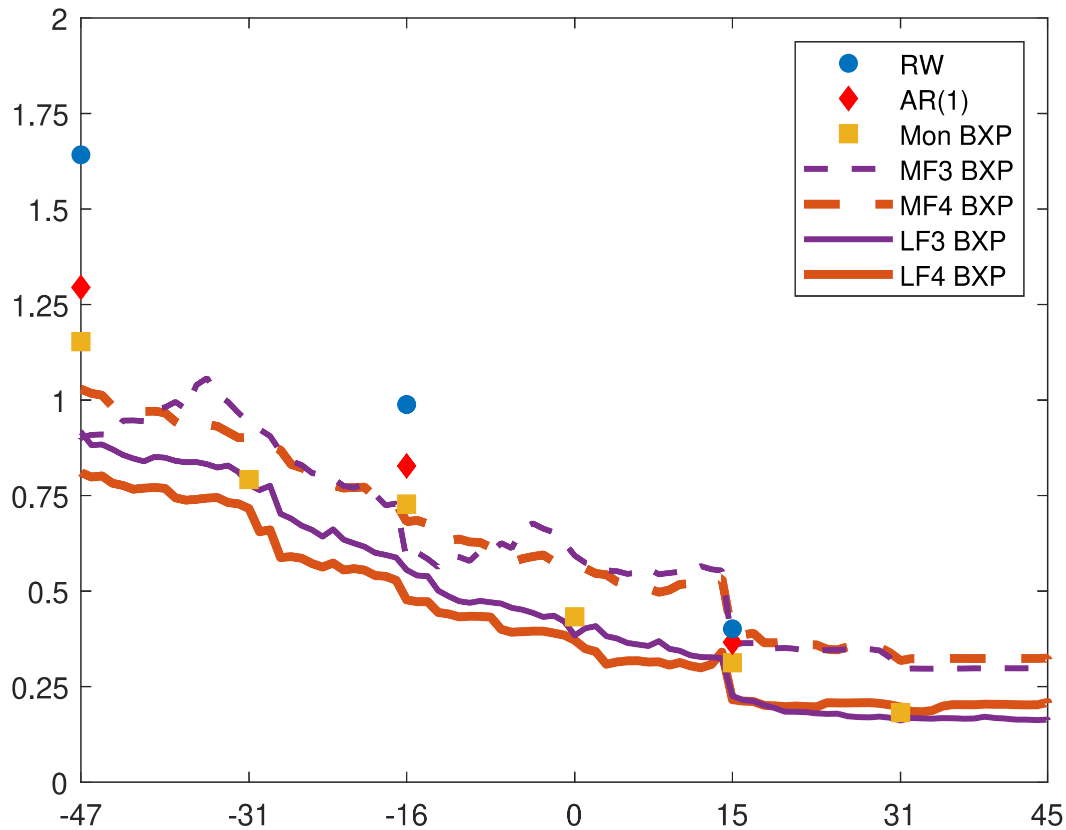
Notes: This figure indicates the flow of data released in a specific month, July 2017, which includes the weekly China Commodity Price Index (CCPI), the 10-day producer goods market price (PGMP), the producer price index (PPI), and the PMI purchasing price index.

Figure 11: Root Mean Squared Forecast Errors: 2011:1-2020:12



Notes: The y-axis reports root mean squared forecast errors (RMSFEs) over the period from January 2011 to December 2020. The x-axis reports the distance in terms of days away from the beginning of the current month, where 0 refers to the last day of the previous month. RW refers to random walk; AR(1) refers to an autoregressive model with lag order one; Mon BXP refers to the best model ex post that includes only monthly variables, i.e., the factor model with 3 factors and 2 lags; MF3 BXP refers to the best mixed-frequency factor model ex post that uses 3 types of data frequencies, i.e., the factor model with 3 factors and 31 lags; and LF3 BXP denotes the best model ex post of the low-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 4 factors and 1 lag.

Figure 12: Root Mean Squared Forecast Errors: 2017:1-2020:12



Notes: The y-axis reports root mean squared forecast errors (RMSFEs) over the period from January 2017 to December 2020. The x-axis reports the distance in terms of days away from the beginning of the current month, where 0 refers to the last day of the previous month. LF4 BXP denotes the best model ex post of the low-frequency dynamic factor model with 4 types of data frequencies, i.e., the factor model with 2 factors and 6 lags; LF3 BXP denotes the best model ex post of the low-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 4 factors and 1 lag; MF4 BXP denotes the best model ex post of the mixed-frequency dynamic factor model with 4 types of data frequencies, i.e., the factor model with 1 factor and 3 lags; MF3 BXP denotes the best model ex post of the mixed-frequency dynamic factor model with 3 types of data frequencies, i.e., the factor model with 3 factors and 31 lags; Mon BXP denotes the best model ex post of the dynamic factor model with only monthly frequency series, i.e., the factor model with 3 factors and 2 lags; and AR(1) and RW denote the autoregressive model with lag order 1 and the random walk model, respectively.

A Supplemental Appendix

Table A.1
Data description.

Name	Source
Monthly	
CN: Producer Price Index (PPI)	Author construction
CN: Producer Price Index: Producer Goods: Mining and Quarrying	Author construction
CN: Producer Price Index: Producer Goods: Raw Material	Author construction
CN: Producer Price Index: Producer Goods: Manufacturing	Author construction
CN: Producer Price Index: Consumer Goods: Food	Author construction
CN: Producer Price Index: Consumer Goods: Clothing	Author construction
CN: Producer Price Index: Consumer Goods: Daily Sundries	Author construction
CN: Producer Price Index: Consumer Goods: Durables	Author construction
CN: PMI: Mfg: Purchasing Price Index	NBS
10-Day	
CN: Producer Goods Price: Market Price: Ferrous Metal: Rebar: 16-25 mm, HRB400E	NBS
CN: Producer Goods Price: Market Price: Ferrous Metal: Wire Rod: 6.5 mm, HPB300	NBS
CN: Producer Goods Price: Market Price: Ferrous Metal: Ordinary Medium Plate: 20 mm, Q235	NBS
CN: Producer Goods Price: Market Price: Ferrous Metal: Ordinary Hot Rolled Sheet: 3 mm, Q235	NBS
CN: Producer Goods Price: Market Price: Ferrous Metal: Seamless Tube: 219*6, 20#	NBS
CN: Producer Goods Price: Market Price: Ferrous Metal: Angle Steel: 5#	NBS
CN: Producer Goods Price: Market Price: Non Ferrous Metal: Electrolytic Copper: 1#	NBS
CN: Producer Goods Price: Market Price: Non Ferrous Metal: Aluminum Ingot: A00	NBS
CN: Producer Goods Price: Market Price: Non Ferrous Metal: Lead Ingot: 1#	NBS
CN: Producer Goods Price: Market Price: Non Ferrous Metal: Zinc Ingot: 0#	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Sulphuric Acid: 98%	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Bake Alkali: Liquid Caustic, 32%	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Methanol: Refine	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Benzene: Industrial	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Styrene: First Class	NBS

Continued on next page

Table A.1 – continued from previous page

Name	Source
CN: Producer Goods Price: Market Price: Petrochemical Product: Polythene: LLDPE, 7042	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Polypropylene: T30S	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: PVC: SG5	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Polybutadiene Rubber: BR9000	NBS
CN: Producer Goods Price: Market Price: Petrochemical Product: Polyester Filament: FDY150D/96F	NBS
CN: Producer Goods Price: Market Price: Petroleum & Natural Gas: LNG	NBS
CN: Producer Goods Price: Market Price: Petroleum & Natural Gas: LPG	NBS
CN: Producer Goods Price: Market Price: Petroleum & Natural Gas: Gasoline: 95#	NBS
CN: Producer Goods Price: Market Price: Petroleum & Natural Gas: Gasoline: 92#	NBS
CN: Producer Goods Price: Market Price: Petroleum & Natural Gas: Diesel Oil: 0#	NBS
CN: Producer Goods Price: Market Price: Petroleum & Natural Gas: Paraffin Wax: 58#	NBS
CN: Producer Goods Price: Market Price: Coal: Anthracite Coal	NBS
CN: Producer Goods Price: Market Price: Coal: 4500 Calorie: Common	NBS
CN: Producer Goods Price: Market Price: Coal: 5000 Calorie: Shanxi Common	NBS
CN: Producer Goods Price: Market Price: Coal: 5500 Calorie: Shanxi High-quality	NBS
CN: Producer Goods Price: Market Price: Coal: 5800 Calorie: Datong High-quality	NBS
CN: Producer Goods Price: Market Price: Coal: Coking Coal: Metallurgical	NBS
CN: Producer Goods Price: Market Price: Construction Material: Common Portland Cement: 42.5 Intensity, Bulk	NBS
CN: Producer Goods Price: Market Price: Construction Material: Float Flat Glass: 4.8/5 mm	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Rice: Round Shaped	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Wheat: Third Class	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Corn: Yellow, Second Class	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Cotton: Ginned Cotton, Third Class	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Live Pig: Hybridize	NBS

Continued on next page

Table A.1 – continued from previous page

Name	Source
CN: Producer Goods Price: Market Price: Agricultural Product: Glycine Max: Soybean	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Soybean Meal: Crude Protein >=43%	NBS
CN: Producer Goods Price: Market Price: Agricultural Product: Peanut: for Oil	NBS
CN: Producer Goods Price: Market Price: Agricultural Input: Urea: Granule	NBS
CN: Producer Goods Price: Market Price: Agricultural Input: Compound Fertilizer: Potassium Sulfate	NBS
CN: Producer Goods Price: Market Price: Agricultural Input: Pesticide: Glyphosate: 95%	NBS
CN: Producer Goods Price: Market Price: Forest Product: Paper Pulp: Bleached Chemical Pulp	NBS
CN: Producer Goods Price: Market Price: Forest Product: Corrugated Paper: High Strength	NBS
Weekly	
CCPI: Aggregate	MINISTRY OF COMMERCE,PRC
CCPI: Energy	MINISTRY OF COMMERCE,PRC
CCPI: Steel	MINISTRY OF COMMERCE,PRC
CCPI: Mineral Production	MINISTRY OF COMMERCE,PRC
CCPI: Nonferrous Metal	MINISTRY OF COMMERCE,PRC
CCPI: Rubber	MINISTRY OF COMMERCE,PRC
CCPI: Agricultural Product	MINISTRY OF COMMERCE,PRC
CCPI: Livestock	MINISTRY OF COMMERCE,PRC
CCPI: Oil&Oilseed	MINISTRY OF COMMERCE,PRC
CCPI: Sugar	MINISTRY OF COMMERCE,PRC
Daily	
CN: Index: CSI 300 Index	China Securities Index Co.,LTD
CN: Interbank Offered Rate: Weighted Avg: 7 Day	National Interbank Funding Center
CN: Bond Yield: Treasury Bond: 3 Month	China Central Depository & Clearing Co., Ltd.
CN: Bond Yield: Treasury Bond: 10 Year	China Central Depository & Clearing Co., Ltd.
CN: CNY Central Parity Rate: USD	China Foreign Exchange Trading Center
Price at Railway Station: 1/3 Coking Coal (V37%, A9.5%, S0.3%, G>65) Produced in Huaibei City, Anhui Province	Steelhome
Market Price: Second-Grade Coke (A<13.0%, S<0.7%, Mt<8%, M25>88%, CSR52-55%) Produced in TangShan City, Hebei Province	Steelhome
Factory Price: High-density Polyethylene HDPE(5000S): Produced by Daqing Petrochemical Company	Zhejiang China Plastics Online Co., Ltd.
Factory Price: Polypropylene PP(T30S): Produced by Dalian Petrochemical Organic Synthesis Co., Ltd.	Zhejiang China Plastics Online Co., Ltd.
Market Price in Beijing City: Hot-rolled Plain Steel Bar (HPB) 300	Wind
Market Price in Qian'an City: Iron Powder Fe Content 66%: Dry Basis Tax Included	Wind

Continued on next page

Table A.1 – continued from previous page

Name	Source
Market Price in Beijing City: Hot Rolled Steel Coil: Q235B:3.0 mm	Wind
Market Price in Beijing City: Channel Steel 16#	Wind
Average Market Price in Shanghai Yangtze River Nonferrous Metals Spot Market: Copper 1#	News report
Average Market Price in Shanghai Yangtze River Nonferrous Metals Spot Market: Aluminum A00	News report
China Cotton Price Index: 328	China National Cotton Exchange