



Contents lists available at ScienceDirect

## International Journal of Forecasting

journal homepage: [www.elsevier.com/locate/ijforecast](http://www.elsevier.com/locate/ijforecast)

# How can big data enhance the timeliness of official statistics? The case of the U.S. consumer price index

Tarek M. Harchaoui<sup>a,b,\*</sup>, Robert V. Janssen<sup>a</sup><sup>a</sup> Department of Global Economics Management, University of Groningen, Netherlands<sup>b</sup> Groningen Growth and Development Centre, University of Groningen, Netherlands

## ARTICLE INFO

**Keywords:**  
Forecasting  
Inflation  
Online prices

## ABSTRACT

The daily consumer price index (CPI) produced by the Billion Prices Project (BPP CPI) offers a glimpse of the direction taken by consumer price inflation in real time. This is in contrast to the official U.S. CPI, which is compiled monthly and released with an average of a three-week delay following the end of the reference month. A recent body of research contended that the movements of online prices are representative of those of offline retail prices, making the BPP CPI a natural candidate for accurately improving the timeliness of the official CPI. We assess the predictive content of the BPP CPI using a variety of MIDAS models that accommodate data sampled at different frequencies. These models generate estimates that remain robust to the variety of time periods considered and, by the standard of the existing literature, contribute to a significant upgrade in the forecast accuracy of official consumer price inflation figures. The paper then sketches the broad implications of BPP CPI for the consumer price statistics maintained by national statistics offices and discusses how the proposed improvement in the timeliness of the official CPI fits in this perspective.

© 2018 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Over recent decades, the Bureau of Labor Statistics (BLS) has made a sustained effort to improve its consumer price index (CPI) program, with the aim of enhancing its credibility (henceforth, official CPI). This flagship program has evolved over the years, stimulated by both new developments in the theory and practice of price indexes and the needs of data users (see Reinsdorf & Triplett, 2009). However, despite the beneficial advances of this program, legitimate concerns about its relevance have been expressed in recent years, mainly suggesting that the massive deployment of information technology in the workplace over the

last twenty years has coincided with a deterioration of the timeliness of such statistics.<sup>1</sup>

Simultaneously, the Billion Prices Project (BPP) began to produce a daily variant of the CPI based on web-scraping of online retail prices on a wide array of products (henceforth, BPP CPI; see Cavallo & Rigobon, 2016). This major breakthrough swept through the users of the U.S. consumer price inflation statistics and has the potential to considerably alter the traditional business model shaped by national statistical offices (NSOs), which features an offline collection of retail prices. Although this virtual real-time CPI is still in the early stages of development, there is abundant anecdotal evidence that it provides policy-makers with a reasonably good “pulse” as to the direction being taken

\* Correspondence to: Faculty of Economics and Business, Global Economics & Management, Nettelbosje 2, 9747 AE Groningen, The Netherlands

E-mail address: [t.m.harchaoui@rug.nl](mailto:t.m.harchaoui@rug.nl) (T.M. Harchaoui).

<sup>1</sup> Between the second half of the 1990s and more recent years, the amount of time after the reference month that the U.S. CPI was published went from nearly two weeks to three weeks.

by inflation in real time.<sup>2</sup> Some more solid evidence in its favour has also started to emerge, with Argentina representing a case in point, where the BPP CPI proved to be a serious alternative to the official CPI after the latter lost its credibility due to political interference (see Cavallo, 2013, and *The Economist*, 2012; see also Cavallo & Bertolotto, Unpublished).

This paper argues that, under an adequate modelling strategy, the daily BPP CPI can be utilized fruitfully to enhance the timeliness of the official CPI reliably. In particular, we assess the predictive information content of the BPP CPI using the mixed data sampling (MIDAS) regression models proposed by Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels, Sinko, and Valkanov (2007). Although we follow a large and growing body of literature emphasizing the benefits of combining data of different frequencies for the forecasting of flagship economic indicators such as consumer price inflation, we are not aware of any other study that has investigated whether the accuracy of consumer price inflation forecasts can be improved by considering the daily BPP CPI. As was emphasized by Cavallo (2017), although most transactions still occur offline in stores, the movements of online prices are representative of retail prices, and thus provide the BPP CPI with a reasonably accurate glimpse of the direction taken by consumer price inflation in real time. Given the importance of forecasting inflation and the considerable attention that the subject has received in the academic literature, it is somewhat surprising that economists have not previously considered the BPP CPI as a natural candidate for anticipating movements in the official CPI.<sup>3</sup>

We assess the forecasting abilities of our MIDAS models through an application to U.S. inflation (all items). The performances of these models are contrasted with that of the AR model, which we regard as a benchmark in accordance with well-established practice (e.g., Bruneau, De Bandt, Flageollet, & Michaux, 2007, Flavin, Panopoulou, & Pantelidis, 2009, Smith, 2015). Our results suggest that the MIDAS models combined with the daily BPP CPI generally provide robust estimates of the time period considered, and contribute to significantly large increases in the forecast accuracy of consumer price inflation. For example, the reduction in the average relative root mean squared forecast errors (RMFSE) for the beta and exponential variants of the MIDAS models hovers around 0.40 for a 12-month horizon forecast—a large improvement by the standard of

the existing literature—and our results also show a statistically significant directional accuracy as high as 0.60.

The early applications of MIDAS models to financial data were later extended to the GDP (see Smith, 2015, and the references therein). Their recent consideration for inflation constitutes further evidence of the relevance of these models. Our paper contributes to this topic that has been gaining in momentum over recent years. Monteforte and Moretti (2013) constitute a recent example of this development. They applied a MIDAS model to financial variables in an effort to extract timely information for forecasting euro inflation in real time, and their results support the view that daily variables contribute to the reduction of forecast errors. Thus, our attempt to exploit daily online prices seems to be a step in the right direction. Another line of research has considered how internet search data can add value to the forecast performances of existing prediction models. Wu and Brynjolfsson (2015) were amongst the first to show that internet queries provide a reliable early signal regarding housing prices and sales long before they change in the marketplace. Similarly, Li, Shang, Wang, and Ma (2015) employed daily consumer prices compiled from Google search data in a MIDAS model for forecasting Chinese CPI, a combination that enhanced the forecast accuracy of official price inflation series considerably. While these internet queries have contributed to the compilation of timely prices, the underlying data are sensitive to the selection of search terms (see Smith, 2016, p. 264, for a discussion).

The daily consumer price inflation index, developed as part of the BPP, uses daily online prices collected by software that scours the websites of online retailers for their prices. As a result, this index constitutes a serious alternative to its competitors based on internet search data. Aparicio and Bertolotto (Unpublished) used the monthly variant of the BPP CPI to forecast the official CPI within a standard VAR model. However, this time aggregation of the BPP CPI comes at the cost of a loss of this potentially relevant high-frequency information, leading to what Foroni, Marcellino, and Schumacher (2015) refer to as “a convolution of the dynamic relationships among the variables”. By using the MIDAS models with the daily CPI BPP, our paper can be viewed as an attempt to relax the restrictions that Aparicio and Bertolotto imposed on their forecast model. While we make no claim of either methodological or conceptual innovation, our paper stresses the importance of the BPP CPI for enhancing the timeliness of the official CPI.

The remainder of the paper is organized as follows. Section 2 sketches the class of MIDAS models utilized in this paper. Section 3 features the source data, the analysis of the descriptive statistics, and the econometric results of the competing models. The quantitative analysis of this section culminates in a robustness check and an analysis of the forecast performances of the alternate models. Section 4 discusses the implications of big data prices for official statistics, and highlights the differences between the short- and medium-run time horizons. The conclusion summarizes the results.

<sup>2</sup> For example, Zumbun (2015), journalist at the Wall Street Journal, recently touted that “The BPP thinks inflation may have turned a sharp corner”, highlighting the early indication that the aggregate demand had been much firmer than officially thought. This is not the first time that the business press has brought to the fore cases of daily internet consumer price changes that often foreshadow subsequent changes in official price indexes. Previously, Surowiecki (2011), journalist at the New Yorker, had revealed that the BPP CPI plunged on the day when Lehman Brothers collapsed in September 2008, as online retailers in America immediately cut prices—an early indication that aggregate demand had weakened. The official CPI did not report this sharp decline until November, when October data were released by the BLS.

<sup>3</sup> For example, none of the many potential uses of big data that were discussed in Charles Bean’s Independent Review of UK Economic Statistics (Bean, 2016) were associated with the enhancement of timeliness.

Download English Version:

<https://daneshyari.com/en/article/7408117>

Download Persian Version:

<https://daneshyari.com/article/7408117>

[Daneshyari.com](https://daneshyari.com)