How can big data enhance the timeliness of official statistics? The case of the US consumer price index

Harchaoui, Tarek M.; Janssen, Robert V.

Published in:
International Journal of Forecasting

DOI:
10.1016/j.ijforecast.2017.12.002

International Journal of Forecasting 34 (2018) 225–234

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: 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 [a,b,*], Robert V. Janssen [a]

[a] Department of Global Economics Management, University of Groningen, Netherlands
[b] 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.

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.[1]

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 (T.M. Harchaoui).

[1] 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.

https://doi.org/10.1016/j.ijforecast.2017.12.002

by inflation in real time.[2] 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.[3]

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 Bertoletto 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.

[2] For example, Zumbrun (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.

[3] 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.

2. The MIDAS models

Under the MIDAS approach, a lower-frequency dependent variable is regressed on a higher-frequency lagged independent variable. The MIDAS model with one regressor forecasting h periods ahead takes the form

$$y_{t+h} = \alpha + \beta B\left(L^{\frac{1}{m}}; \theta\right) x_t^{(m)} + \xi_{t+h}, \tag{1}$$

where y_t is the regressand and $x_t^{(m)}$ is the regressor, sampled at frequency m. If the regressand and the regressor are sampled at the same frequency, then $m = 1$, whereas if the regressand is sampled monthly while the regressor is sampled daily, $m = 30$. The polynomial lag operator $B\left(L^{\frac{1}{m}}; \theta\right)$ works on $x_t^{(m)}$, weighting each lagged observation, where

$$B\left(L^{\frac{1}{m}}; \theta\right) = \sum_{k=1}^{K} B\left(k; \boldsymbol{\Theta}\right) L^{\frac{k}{m}}, \tag{2}$$

while k is the number of lags and $L^{\frac{1}{m}}$ is a lag operator, such that $L^{\frac{1}{m}} x_t^{(m)} = x_{t-1/m}^{(m)}$. For $B\left(k; \boldsymbol{\Theta}\right)$, where $\boldsymbol{\Theta}$ is a vector of parameters to be estimated, we use three polynomial weighting functions (see Armesto, Engemann, & Owyang, 2010, for a review).

The first polynomial weighting function that we use is the Almon lag (Almon, 1965), where the weight on each lag k is calculated as $B\left(k; \boldsymbol{\Theta}\right) = \sum_{q=0}^{Q} \theta_q k^q$, where Q denotes the order of the polynomial. The second is the beta polynomial structured around two parameters, and is defined as follows:

$$B\left(k; \boldsymbol{\Theta}\right) = \frac{f(\frac{k}{K}, \theta_1, \theta_2)}{\sum_{k=1}^{K} f(\frac{k}{K}, \theta_1, \theta_2)}, \tag{3}$$

where:

$$f\left(\frac{k}{K}, \theta_1, \theta_2\right) = \frac{\left(\frac{k}{K}\right)^{\theta_1 - 1}\left(1 - \frac{k}{K}\right)^{\theta_2 - 1} \Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}, \tag{4}$$

and

$$\Gamma\left(\theta_p\right) = \int_0^\infty e^{-\left(\frac{k}{K}\right)} \left(\frac{k}{K}\right)^{\theta_p - 1} d\left(\frac{k}{K}\right), \tag{5}$$

with $\Gamma\left(\theta_p\right)$ representing the standard gamma function. The third polynomial is the exponential structure introduced by Ghysels et al. (2007), whom we follow by assuming the functional form of two parameters, $\boldsymbol{\Theta} = [\theta_1, \theta_2]$:

$$B\left(k; \boldsymbol{\Theta}\right) = \frac{e^{\left(\theta_1 k^1 + \theta_2 k^2\right)}}{\sum_{k=1}^{m} e^{\left(\theta_1 k^1 + \theta_2 k^2\right)}}.$$

These three variants of the MIDAS model are at the core of our empirical implementation.

3. Quantitative analysis

3.1. The source data

The daily CPI is the result of the BPP initiative launched by the MIT for research activities (see Cavallo & Rigobon, 2016), an academic initiative that later turned into a commercial activity structured within PriceStats, an entity that is responsible for collecting, assembling, constructing, and analyzing the daily CPI for the U.S. and a growing number of other countries. An important feature of this initiative is its attempt to use information technology effectively in order to harness the massive amount of information that is available online. Typically, web-scraping algorithms are utilized to track, on a daily basis, a panel of 30,000 distinct products that are identified uniquely and are available from a wide range of retailers throughout the U.S. The sample of products is selected carefully in an attempt to prevent an overrepresentation of goods that are sold abundantly online, such as electronics, cosmetics, and books. The attributes of these products are summarized by their unique IDs. This feature facilitates the rotation of the sample that, in turn, accommodates product turnover. The 15 million online prices that are collected on a daily basis from more than a thousand retailers are then utilized, along with the official category CPI weights, to construct an aggregate price index of consumer's inflation for the U.S. and a number of other major economies.

The official CPI, on the other hand, is based on data from 22,000 outlets across 88 geographic areas.[4] The BLS compiles the information on food and beverages, housing, apparel, transportation, medical care, recreation, education and communications, and other goods and services. Prices of food and energy are collected monthly, while those of other items are collected bimonthly. The sample selection features a two-step process where outlets are selected probabilistically using household point-of-purchase surveys, then items within outlets are selected based on their relative sales. The prices of these items are collected using a detailed checklist of their attributes in order to ensure that the same item is priced in consecutive months ("pure" pricing). In the event that the item that was meant to be priced has been discontinued, the BLS' procedure is to begin pricing a closely related item at the same outlet, with a correction based on historic prices. The CPI measures the (official) average (pure) price change for a set of consumers' goods and services purchased by all urban consumers (CPI-U), representing approximately 89% of the total population; urban wage earners and clerical workers (CPI-W), representing 28% of the population, are also covered by the CPI. Sales and excise taxes, as well as using fees, are included in the measure of consumers.

The publicly available version of the BPP CPI runs from July 1, 2008, to July 31, 2015, implying that we have to curtail the official CPI data availability to match this time period. While the official CPI and the BPP CPI offer some scope for complementarity, particularly in the areas of timeliness, there are several differences that set them apart which are hard to settle within the scope of this paper. For example, Pricestats.com publishes the BPP CPI with a only three-day lag with respect of the reference month, which implies that it can be used effectively to predict the official CPI that is available with an average of a three-week delay from the reference month. Of the many differences, that of coverage is the most acute. For example, the BPP CPI

4 See https://www.bls.gov/opub/hom/pdf/homch17.pdf.

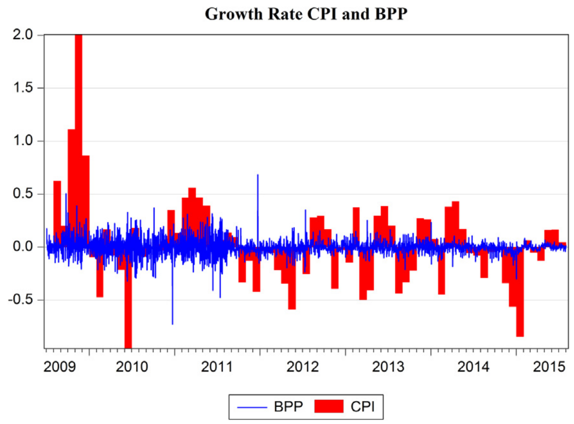

Fig. 1. Growth rates of official CPI and BPP CPI (daily and monthly variation in %).

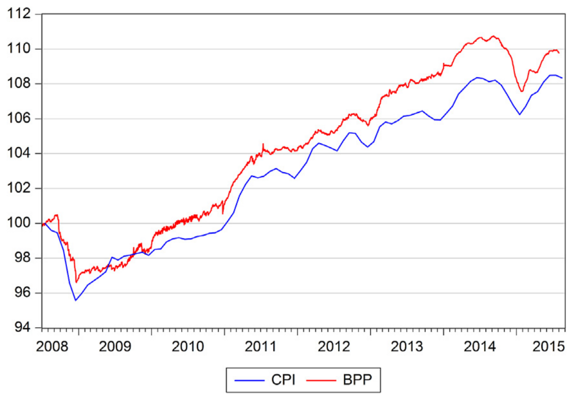

Fig. 2. Non-seasonally adjusted official CPI and BPP CPI (2008 = 100).

does not account for either medical services or retail fuel prices, despite their ready availability online. Given that fuel prices are highly volatile, their exclusion is expected to lead to wide discrepancies between the two measures of inflation. In the absence of a better solution, we leave the coverage unchanged, which Cavallo and Rigobon (2016) found did not have a fundamental impact on the consumer price dynamics.[5]

Another difference is the lack of a seasonally adjusted BPP CPI, unlike its official counterpart. These differences may generate different patterns between the two types of CPIs. One possible way of alleviating the effect of seasonality would be to bring the official CPI to its lowest common denominator, represented by the CPI excluding energy, which is not seasonally-adjusted. However, while this option is appealing, it is not feasible under the U.S. CPI program, which excludes both food and energy. If anything, this option will make things worse. Thus, we instead endeavour to make the movements of the two CPIs comparable by constructing the 12-month variation with respect to the same day or the same month, a practice that is used widely by official statistics programs.

Fig. 1 displays the daily and monthly variation in the BPP and official CPIs (raw data). The daily BPP CPI fluctuates within the monthly official CPI series, conveying potentially valuable information. MIDAS represents a class of models that offer the attractive features of relating variables sampled at different frequencies directly without the loss of high-frequency information and accommodating the dynamic patterns of the data. However, before getting into those models, an analysis of some broad trends in the two source datasets is warranted.

3.2. Trend analysis

The raw data versions of the two CPIs reported in Fig. 2 offer some similar broad long-term patterns, along with some important medium-term differences. The level reported by the BPP CPI is higher than that reported by its official counterpart, reflecting the coverage of services which tend to dampen volatility and growth. The official CPI also displays more pronounced fluctuations, due to the inclusion of fuel components.

Fig. 3 indicates the variation in the two indexes, which report broadly similar patterns for some episodes but not for others, albeit without reporting conflicting directions of movement. This is confirmed by the order of magnitude of the rolling correlation coefficient, which ranges from −0.63 to 0.99, depending on the episode. Again, given the difference in coverage between the two indexes, the broad similarity in their patterns speaks favorably of using the BPP CPI to generate a more current official CPI. Another way of ascertaining the extent to which the official CPI and its BPP counterpart follow the same pattern is to discuss the underlying descriptive statistics. The results reported in Table 1 confirm some earlier findings regarding the similarities in the broad patterns, with the percentage point differences between the two indexes being small, though with the BPP CPI showing a tendency to report a volatility that was slightly in excess of that of its official counterpart.

Table 1
Descriptive statistics: average annual growth rate (%).

	CPI (1)	BPP (2)	(1) *minus* (2)
Mean	1.6	1.8	−0.2
Median	1.7	2.2	−0.6
Maximum	3.8	4.2	−0.4
Minimum	−2.1	−2.7	0.6
Std. Dev.	1.2	1.5	−0.3
Observations	74	2223	

Note: The calculations are based on the twelve-month variation.

4. Econometric implementation

4.1. Baseline estimates and robustness checks

We estimate the MIDAS model under the three polynomial weighting functions: the Almon polynomial lag, the beta polynomial lag, and the exponential polynomial lag. While these three methods share similar estimation

[5] In particular, they found that not only are online and offline prices identical in 72% of the cases, but online prices can also closely approximate the official CPI.

Table 2
Regression results of the MIDAS Almon weighting with alternate polynomial degrees: 2009:07–2015:07.

Polynomial degree	$p = 1$	$p = 2$	$p = 3$	$p = 4$
α	0.0059	0.0018	−0.0017	0.0018
	(0.8859)	(0.9633)	(0.9648)	(0.9634)
β	0.2912***	0.4156***	0.4070***	0.4145***
	(0.0091)	(0.0002)	(0.0004)	(0.0003)
Θ_1	0.3532***	1.0461***	0.8519**	0.3921
	(0.0005)	(0.0000)	(0.0485)	(0.4861)
Θ_2		−0.0493***	−0.0130	0.1479
		(0.0012)	(0.8492)	(0.3141)
Θ_3			−0.0012	−0.0138
			(0.5908)	(0.1866)
Θ_4				0.0003
				(0.2158)
Adj. R^2	0.3410	0.4377	0.4402	0.4533

Note: Standard errors are given in parentheses.*,** and*** indicate significance at the 10%, 5% and 1% levels, respectively.

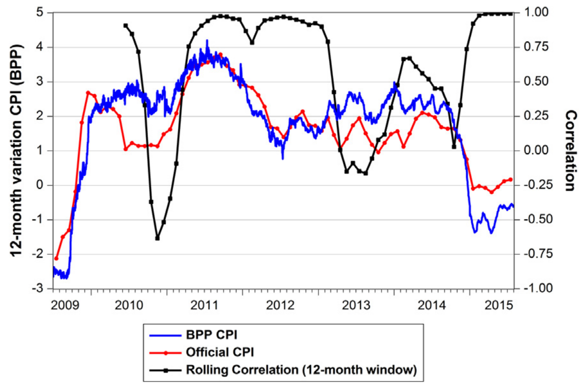

Fig. 3. Rolling correlation and 12-month variation of the CPI (BPP).

Table 3
Regression results of the different MIDAS weighting methods.

	Almon with $p = 2$	Exponential	Beta
α	0.0018	0.0038	0.0009
	(0.9633)	(0,9213)	(0.8182)
β	0.4156***	0.3805***	0.3467***
	(0.0002)	(0.0001)	(0.0008)
γ		11.1195***	11.8613***
		(0.0000)	(0.0000)
Θ_1	1.0461***	0.4508*	1.0557***
	(0,0000)	(0.0775)	(0.0000)
Θ_2	−0.0493***	−0.0372*	3.4076***
	(0.0012)	(0.0635)	(0.0021)
Adj. R^2	0.4377	0.4643	0.4420

Note: Standard errors are given in parentheses.*,** and*** indicate significance at the 10%, 5% and 1% levels, respectively.

methods, represented by non-linear least squares, there are important differences between them. First, for the Almon polynomial weighting function, we estimate up to a degree four of the vector of high frequency coefficients, $\mathbf{\Theta}$, along with the intercept α and the autoregressive parameter β. Second, the beta and the exponential polynomial lag require the estimation of four parameters: $\theta_1$and θ_2, which structure the shape of the weighting function, the intercept α and the slope β.

The models have been estimated using data from July 1, 2009, to July 31, 2015. The use of the Almon polynomial weighting modelling strategy requires the order of the polynomial degree to be determined. The Akaike information criterion. along with the significance of the parameters, led to the selection of polynomial degrees ranging from one to four. The corresponding results are reported in Table 2. The combination of the adjusted-R^2 and the significance of the parameters suggests that the model with a polynomial structure of order two performs best, so it will be retained for the forecasting exercise.

Table 3 reports the results of the MIDAS models with both exponential and beta weighting structures. The Almon weighting structure with an order of two has also been repeated for the sake of completeness. The results suggest that it is difficult to discriminate between the three models based on the significance of their parameters, as they all turn out to be significant, though with varying degrees of uncertainty. In this particular respect, the Almon and beta models are slightly superior to the exponential. However, when considering the proportion of the variance that each of these models explains, the exponential and the beta emerge, though the order of magnitude difference is not large. Given that there are arguments both for and against for each model, all of them have been retained for the forecasting exercise.

We now confirm the validity of our approach by checking the robustness of our results through the consideration of a different sample period. The robustness check is assessed in terms of the stability of the parameters. The sample period is now split into two sub-periods, pre- and post-12/2012, defined on the basis of the quantitative easing employed by The Fed to stimulate the economy by large-scale purchases of assets in exchange for cash injections. Of the several rounds of quantitative easing that were initiated in 12/2008 and officially brought to a close in 10/2014, the first three were the largest in terms of the scale of asset purchases, and thus define the split of our sample period.[6]

6 The first round of quantitative easing (12/2008–06/2010) involved \$1.25 trillion in mortgage-based securities and \$300 billion in long-term treasuries. The second round (11/2010–06/2011) involved \$600 billion in long-term treasuries. The third, called Operation Twist (09/2011–12/2012), involved \$400 billion in long-term treasuries. The last round

Table 4
Descriptive statistics for the sub-period 2009:07–2012:12 and 2013:01–2015:07: average annual growth rate (%).

	Official CPI (1)		BPP CPI (2)		(1) minus (2)	
	2009:07–2012:12	2013:01–2015:07	2009:07–2012:12	2013:01–2015:07	2009:07–2012:12	2013:01–2015:07
Mean	1.92	1.15	2.01	1.44	−0.08	−0.29
Median	2.04	1.41	2.32	2.12	−0.28	−0.71
Maximum	3.80	2.10	4.21	2.99	−0.41	−0.89
Minimum	−2.12	−0.20	−2.70	−1.39	0.58	1.19
St. Dev.	1.33	0.74	1.53	1.37	−0.20	−0.62
N	42	32	1280	943		

Table 5
Regression results of the MIDAS Almon weighting with alternate polynomial degrees: different sub-periods.

	$p = 1$		$p = 2$		$p = 3$		$p = 4$	
	2009:07–2012:12	2013:01–2015:07	2009:07–2012:12	2013:01–2015:07	2009:07–2012:12	2013:01–2015:07	2009:07–2012:12	2013:01–2015:07
α	0.0029	−0.0222	−0.0042	−0.0323	−0.0147	−0.0319	0.0009	−0.0341
	(0.9641)	(0.6624)	(0.9437)	(0.4628)	(0.8156)	(0.4772)	(0.9887)	(0.4578)
β	0.3881***	−0.1140	0.5430***	−0.0562	0.5348***	−0.0604	0.5286***	−0.0554
	(0.0071)	(0.5576)	(0.0005)	(0.7379)	(0.0007)	(0.7287)	(0.0008)	(0.7548)
Θ_1	0.3821**	0.4160***	1.1527***	1.2062***	0.8646	1.1480**	0.2014	0.9622
	(0.0111)	(0.0028)	(0.0013)	(0.0001)	(0.1705)	(0.0352)	(0.8200)	(0.1532)
Θ_2			−0.0538**	−0.0560***	−0.0002	−0.0449	0.2046	0.0370
			(0.0153)	(0.0028)	(0.9983)	(0.6053)	(0.3523)	(0.8493)
Θ_3					−0.0017	−0.0004	−0.0175	−0.0071
					(0.5831)	(0.8959)	(0.2577)	(0.6261)
Θ_4							0.0003	0.0001
							(0.2964)	(0.6372)
adj. R^2	0.3916	0.3012	0.4845	0.5005	0.4890	0.5008	0.5054	0.5053

Note: Standard errors are given in parentheses.*,** and*** indicate significance at the 10%, 5% and 1% levels, respectively.

We begin with the set of descriptive statistics reported in Table 4 that corroborate the earlier findings, namely that the growth rate reported by the official CPI is smaller than that of the BPP CPI, but with no difference in regard to the direction of the trend. While this pattern remains consistent across the two sub-periods, the difference that emerged during the 2013:01–2015:07 sub-period is larger than that of the earlier sub-period.

Comforted by the notion that the split between the two sub-periods is supported by some differences in the growth path, we re-estimated the models discussed above for each of the sub-periods considered. The counterparts of Tables 2 and 3 are Tables 5 and 6, respectively. Of the various polynomial orders considered for the Almon weighting structure, the one with an order of two outperforms the others in terms of the significance of the parameters across the two periods, a result that is consistent with that obtained under the baseline model estimated for the 2009:07–2015:07 period. The results of the preferred Almon model are reported in Table 6 alongside those of the models with the exponential and beta structures. The results indicate that the models with the Almon and beta structures emerge as the ones with the most sensible results. Their parameters are consistent from one period to the next and remain significant, though the beta explains a somewhat larger proportion of the variance. All things considered, this set of results highlighting the preference for the Almon and beta models corroborates those reached earlier with the full sample period.

(09/2012–10/2014) had only \$85 billion in long-term treasuries and \$40 billion monthly in MBS. Thus, the first three account for 95% of the asset purchases in the entire cycle.

4.2. Forecast results

Given the robustness of the econometric results, we now move to the forecasting part of this paper. The models reported in Table 3 are re-estimated for the 2009:07–2014:06 period, and our simulated out-of-sample forecasting experiment is conducted using a recursive methodology. The out-of-sample forecast period is 2014:07 to 2015:07, giving a ratio of out-of-sample to in-sample observations of 0.18, and features an inflation decline in the early period followed by a steady pick-up. We re-estimate all of the candidate models at each step by adding one observation at a time, then generate h-step-ahead forecasts for months 1–12 and calculate the corresponding RMSFEs. The forecast results obtained from the MIDAS models are then compared to those from the autoregressive model with one lag that is considered as the benchmark model. The same estimation window and forecasting method are employed for the benchmark model as for the MIDAS. RMSFE ratios below one indicate that the MIDAS model in question has a lower forecast error than its AR(1) counterpart.

The forecast results are reported in Fig. 4, which displays the variations of the inflation forecasts and the realized inflation. Apart from a few episodes (e.g., 2014:07, 2014:10, and 2015:02), it appears that, overall, our models reliably replicate the movements of the official CPI. Table 7 reports the RMSFE ratios in order to gauge the forecast accuracy of the MIDAS models; we also computed the Diebold–Mariano test so as to assess whether this improvement in the forecasts (Diebold & Mariano, 1995) is statistically significant. The results show that, the majority of the MIDAS

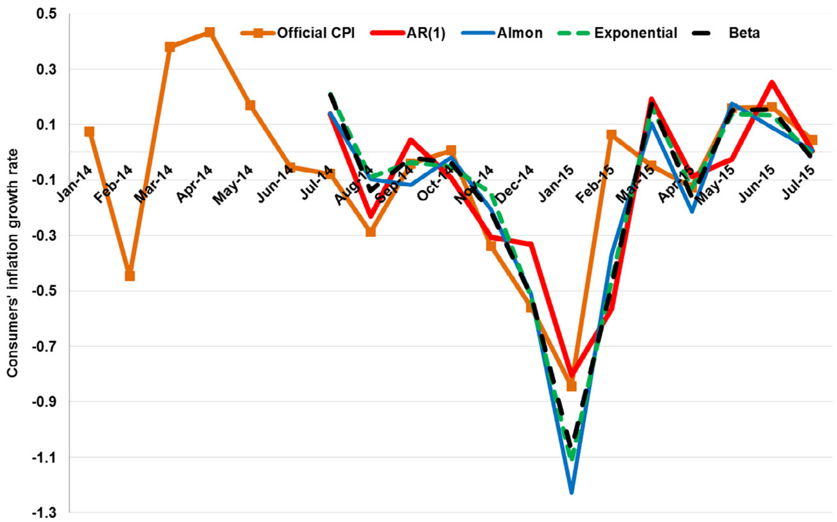

Fig. 4. Forecasts of U.S. official consumers' inflation based on alternate models.

Table 6
Regression results of the different MIDAS weighting methods: different sub-periods.

	Almon with $p = 2$		Exponential		Beta	
	2009:7–2012:12	2013:01–2015:7	2009:7–2012:12	2013:01–2015:7	2009:7–2012:12	2013:01–2015:7
α	−0.003	−0.038	0.006	−0.037	0.007	−0.032
	(0.981)	(0.282)	(0.915)	(0.421)	(0.905)	(0.472)
β	0.691***	−0.034	0.499***	−0.077	0.460***	−0.091
	(0.002)	(0.807)	(0.000)	(0.635)	(0.001)	(0.573)
γ			11.626***	12.669***	12.596***	13.981***
			(0.000)	(0.000)	(0.001)	(0.000)
Θ_1	1.490***	0.944***	0.504	0.268	1.041***	1.059***
	(0.005)	(0.000)	(0.123)	(0.467)	(0.000)	(0.000)
Θ_2	−0.082**	−0.035**	−0.039*	−0.028	3.043**	4.919**
	(0.014)	(0.016)	(0.096)	(0.410)	(0.019)	(0.026)
Adj. R^2	0.519	0.389	0.525	0.495	0.485	0.517

Note: Standard errors are given in parentheses.
* Indicate significance at the 10% level.
** Indicate significance at the 5% level.
*** Indicate significance at the 1% level.

models outperform the benchmark model for nine of the 12 months that were considered. The poor performances of the MIDAS models occur particularly at the beginning and end of the out-of-sample forecast periods. However, although the MIDAS models perform less satisfactorily for these months, they nonetheless provide similar directions to the change reported by the official CPI. With the exclusion of these "outliers", the average RMSFE favors the beta model, followed by the exponential model (0.397 vs. 0.418). In contrast, while the Almon model still outperforms the benchmark model, it lags behind these two other MIDAS models considerably (0.841).

The relatively poor performance of the Almon polynomial is due to the fact that it is sensitive to the number of lags that are considered, which may vary depending on the time horizon. Moreover, the Almon model performs erratically, generally outperforming the AR(1) model but underperforming in other instances, a result that arises from the fact that we considered the same number of lags regardless of the time horizon. In contrast, the relative performances of the beta and exponential models are more consistent across time horizons, though with the former showing a more favourable performance, particularly towards the end of the period.

At this stage, it is useful to contrast the results reported in Table 7 with other competing figures, such as those provided by Monteforte and Moretti (2013), based on the MIDAS model. Given the differences with respect to the time period and the nature of the variables used for the real-time forecasts of inflation, no definitive reconciliation with our results is possible, making the only feasible comparison a qualitative one, which is entertained as a crude way of providing some perspective on our results. For a one-month forecast, their average forecast accuracy ratio of the MIDAS model to the AR(1) model is about 0.88, which is considerably larger than ours. Thus, our forecast accuracy improvements are large by the standard of the literature on the exploitation of MIDAS models for inflation forecasting, possibly due to the use of the BPP CPI.

We also report the directional accuracy of the forecasts in Table 8, in the form of the success ratio. The statistical

Table 7
Out-of-sample nowcast testing.

	2014:07	2014:08	2014:09	2014:10	2014:11	2014:12	2015:01	2015:02	2015:03	2015:04	2015:05	2015:06	2015:07	Average
Almon	1.006	3.343	0.909	0.252*	4.374	0.213*	9.983	0.687*	0.640*	2.208	0.090*	0.831*	1.770	0.845
Exp.	1.324	3.460	0.063*	0.594*	6.383	0.149*	7.088	0.843*	0.879*	0.061*	0.112*	0.333*	0.726*	0.418
Beta	1.310	2.628	0.255*	0.043*	4.126	0.182*	5.982	0.878*	0.935	0.904	0.040*	0.108*	0.230*	0.397

Notes: The table shows RMSFE-MIDAS/RMSFE-AR(1) values for the different horizons. Values under one indicate that the MIDAS results are better than the AR(1) results. The relative accuracies of the individual and factor models are assessed by means of the root mean square forecasting error (RMSFE), while a classic Diebold and Mariano (1995) test for comparative predictive accuracy is employed to assess the statistical significance of any improvement (or deterioration) in the RMSFE relative to the benchmark model.

Table 8
Directional forecast accuracy statistics: success ratios.

	2014:07	2014:08	2014:09	2014:10	2014:11	2014:12	2015:01	2015:02	2015:03	2015:04	2015:05	2015:06	2015:07	Average
Almon	0.297	0.341	0.421	0.385	0.511*	0.614*	0.512*	0.381	0.335	0.514*	0.598*	0.507*	0.438	0.543
Exp.	0.336	0.354	0.621*	0.313	0.412	0.667*	0.524*	0.354	0.321	0.684*	0.621*	0.632*	0.325	0.625
Beta	0.328	0.387	0.534*	0.346	0.526*	0.608*	0.548*	0.341	0.302	0.605*	0.698*	0.678*	0.304	0.600
AR(1)	0.284	0.541*	0.289	0.279	0.641*	0.489	0.605*	0.258	0.269	0.536*	0.361	0.513*	0.446	0.567

Notes: The benchmark model is the monthly no-change forecast.
* Indicates a statistically significant improvement in directional accuracy according to the Pesaran–Timmermann test relative to the no-change forecast.

significance of any gains in directional accuracy is evaluated using the test of Pesaran and Timmermann (2009). This indicator is defined as the proportion of times that the MIDAS models predict correctly whether the inflation is increasing or declining. Under the null hypothesis of no directional accuracy, the success ratio is expected to be equal to 0.5, while higher values suggest that a monthly forecasting model has a higher and more statistically significant directional accuracy. Our results suggest that the exponential and the beta models generate higher average values of the success ratio (0.62 and 0.60, respectively) than the Almon model (0.54). While the first two track the direction taken by the official consumer price inflation reasonably well, the performance of the Almon model lags slightly behind that of the benchmark model (0.56).

5. BPP CPI and official statistics: implications in the short- and medium-run

What does the future hold for the official CPI program? Does the BPP CPI represent a complement to the official program, as the main theme of our paper seems to argue? Or is there any prospect of the BPP CPI representing a genuine way forward, and eventually replacing the present business model? To answer these questions, it may be instructive to go back in time to track its progress and to assess whether there are any indications that a new business model is emerging and will gradually supersede the previous one, inherited from the beginning of the 20th century.

In the majority of developed nations, the CPI program has a history more than a century long, making it the longest-standing flagship economic indicator, just behind the Census of Population. The current version of the CPI departs significantly from its first edition, which was produced around the time of WWI and designed to protect the working class from the increase in prices that was driven by the first world conflict. The commodity coverage has since expanded from basic food to services, including high-tech products. The information used for weighting the components of the index has also evolved considerably. The set of rudimentary, highly subjective and unchained weights has given way to changing and more reliable weights, made possible by the reliance on ongoing and sophisticated surveys of household spending. Although the spectrum of commodities has expanded over time, not every price of every single type of good or service purchased by the CPI population group is collected. NSOs generally rely on judgmental samples where representative sets of goods and services are selected for regular pricing. Similarly, it is not practical to observe the prices of the selected goods and services at all retail outlets that sell these items to the CPI population group. Again, NSOs use judgmental sampling to select a representative sample of outlets for pricing the selected items in each collection cycle.

Just as both the product spectrum and their corresponding weights have been upgraded, so the collection modes employed by NSOs have evolved from personal visits to the store by trained staff to the gradual use of third-party data, ranging from scanner data to online prices. The information and communication technology revolution that swept over developed nations some twenty years ago altered both firm and consumer behaviours. At the same time, this technological breakthrough made new source data available in areas that remained inaccessible to long-standing administrative data collection techniques. Scanner data collected by retailers as part of their ongoing business are one example of such new source data. Along with the proliferation of modern store formats, there was also a surge in online sales platforms that facilitated online pricing in areas such as air travel and audio-visual and computing equipment.

For most periods, these new source data have complemented the old business model based on ongoing surveys nicely. They have helped to address important gaps in a cost-effective manner and have reduced the respondent burden of pricing through personal visits. As the cost of storing, sharing, and analysing data has decreased, economic activity has become increasingly digital, and the internet has transformed retailing business further. This development has led to the online availability of a large and increasing share of consumer goods, which has also given rise to an unprecedented volume of high frequency retail price data. With the development of "web scraping" technology, the BPP initiative has proven that such microdata retail prices can be collected remotely at a considerably lower cost than traditional offline collection (Cavallo & Rigobon, 2016).

From an academic experiment, the BPP is increasingly turning into a serious alternative to the traditional NSO business model in the area of consumer price statistics. For example, Cavallo (in press) showed that, except for the spectrum of products and retailers covered, "scraped data" outperform both the traditional CPI sample data and scanner data in a number of areas, such as frequency, real-time availability, sample size, product details (size, brand, sale), comparability across countries, and uncensored price spells. While these results constitute a welcome development, doubts remain as to the accuracy of online prices, which arguably do not meet statistical standards such as representativeness (see Landefeld, 2014, for a discussion on this issue).

One way to address this problem of representativeness can be to "benchmark" online prices to their offline counterparts, which is where the two studies conducted by Cavallo (2013, 2017) come into play. His 2013 study focussed on several Latin American economies, where indices constructed from online prices provide a reasonably good approximation of both the level and movements of the official inflation rates for the majority of countries considered (namely Brazil, Chile, Columbia, and Venezuela). In contrast, the huge discrepancy identified for Argentina has proven that online prices represent an effective tool for data confrontation, which guarantees quality assurance. Cavallo's 2017 study, which constitutes a large-scale extension of the one conducted in 2013, exploited a panel of retail prices from large multi-channel retailers (i.e., who maintain both online and physical stores), covering ten countries over 16 consecutive months. This study provided compelling evidence regarding the representativeness of online prices, despite the fact that an overwhelming share of transactions are still performed offline. For example, nearly $\frac{3}{4}$ of prices are identical across all matched observations. Although the extent of the synchronization of these prices is limited, their variations display similar frequencies and sizes.

Overall, the BPP CPI is constantly proving that it represents the new business model for the future compilation of inflation statistics in a cost-effective manner. It represents a census of online prices collected in a continuous mode rather than a statistical sample featuring data collection at infrequent intervals. Combined with the electronic transmission of prices from point of collection to point of processing, this continuous flow of prices enhances the timeliness and accuracy of the information. Despite what might still be regarded as partial coverage, these online prices "replicate" the main patterns of offline prices quite well, suggesting a reasonable degree of reliability. With this new landscape as a backdrop, the real question is where the future of official inflation statistics stands. Clearly, in the medium-term, the BPP CPI is paving the way to a point where the new business model will supersede the old one, given that the latter stands in stark contrast to the new mantra of the retail business—faster, better, and cheaper. While the old business model still remains at the core of the majority of NSOs, some of them, primarily the UK Office for National Statistics (ONS) and Statistics New Zealand (SNZ), have been preparing actively to switch to a "big data" version of the CPI (see Breton et al., 2016, and Bentley & Krsinich, 2017). While this transition unavoidably takes time, NSOs can still enhance their timeliness by complementing their official CPI with the BPP CPI using the sort of framework exploited in the present paper.

6. Concluding remarks

The unprecedented volume of consumer prices available online, combined with the development of "web scraping" technology that records such prices, is altering the traditional business model for the compilation of consumer price inflation. Under this new environment, the real question is what the future holds for the official CPI programs. The new business model that has manifested itself with the rise of the BPP is effectively proving that the web is turning into a serious alternative platform for the collection of representative prices in a more timely and cost effective manner than the business model that features physical stores. It is clear that this development is here to stay; the only question is how quickly official statistics can adapt to this new business model. While organizations such as SNZ and ONS have increased their level of preparedness, the traditional business model is still the norm for other NSOs for the foreseeable future. This in turn implies that little progress can be expected in the meantime on cost effectiveness and timeliness in the compilation of official consumer price inflation.

While it is difficult to think of any other way of curbing costs other than the use of online pricing, the issue of timeliness is not as hopeless as it may first appear, which is where our paper comes into play. We contend that, under a suitable modelling strategy, the daily BPP CPI can be utilized fruitfully to obtain a more reliable estimate of a timely official CPI. We investigate the predictive content of the BPP CPI through the use of a variety of MIDAS regressions to circumvent the mixed frequency problem and to summarize its information content efficiently. Our results suggest that the combination of the MIDAS models with the daily BPP CPI gives rise to results that remain broadly robust across time periods. We perform monthly out-of-sample official CPI forecasts and find that using BPP CPI in the MIDAS models with a beta and exponential polynomial weighting function reduces the RMSFE significantly compared to the AR baseline model. For the beta and exponential versions of the MIDAS models, the reduction in the average forecast errors is larger than what is typically found in the literature. For example, for a 12-month time horizon, the reduction in the average error relative to the benchmark model was around 0.40, while their directional accuracy was significant and as high as 0.60.

Acknowledgments

We are indebted to Jan Jacobs, the editor (George Kapetanios), and three anonymous referees for valuable comments made on an earlier draft. The usual caveats apply.

References

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, *33*(1), 178–196.

Aparicio, D., & Bertolotto, M. (2016). *Forecasting inflation with online prices*. MIT, (Unpublished) paper.

Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, *92*(6), 521–536.

Bean, C. (2016). Independent review of UK economic statistics. London: UK.

Bentley, A., & Krsinich, F. (2017). Towards a big data CPI for New Zealand. Paper presented at the Ottawa Group 2017, Eltville, Germany.

Breton, R., Flower, T., Mayhew, M., Metcalfe, E., Milliken, N., & Payne, C. et al., (2016). Research indices using web scraped data: May 2016 Update. Office for National Statistics UK., May 23.

Bruneau, C., De Bandt, O., Flageollet, A., & Michaux, E. (2007). Forecasting inflation using economic indicators: the case of France. *Journal of Forecasting*, *26*(1), 1–22.

Cavallo, A. (2013). Online and official price indexes: measuring Argentina's inflation. *Journal of Monetary Economics*, *60*(2), 152–165.

Cavallo, A. (2017). Are online and offline prices similar? Evidence from large multi-channel retailers. *The American Economic Review*, *107*(1), 283–303.

Cavallo, A. Scraped data and sticky prices. *Review of Economics and Statistics* (in press).

Cavallo, A., & Bertolotto, M. (2016). *Filling the gap in Argentina's inflation data*. MIT, (Unpublished) paper.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: using online prices for measurement and research. *Journal of Economic Perspectives*, *30*(2), 151–178.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*, 253–263.

Flavin, T., Panopoulou, E., & Pantelidis, T. (2009). Forecasting growth and inflation in an enlarged Euro area. *Journal of Forecasting*, *28*(5), 405–425.

Foroni, C., Marcellino, M., & Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of Royal Statistical Society, Series A*, *178*, 57–82.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: mixed data sampling regression models. Working paper.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: further results and new directions. *Econometric Reviews*, *26*(1), 53–90.

Landefeld, S. (2014). Uses of big data for official statistics: privacy, incentives, statistical challenges, and other issues. In *International conference on big data for official statistics*, Beijing, China, October.

Li, X., Shang, W., Wang, S., & Ma, J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data. *Electronic Commerce Research and Applications*, *14*, 112–125.

Monteforte, L., & Moretti, G. (2013). Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, *32*(1), 51–61.

Pesaran, M. H., & Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, *104*, 325–337.

Reinsdorf, M., & Triplett, J. E. (2009). A review of reviews: ninety years of professional thinking about the consumer price index. In W. E. Diewert, J. S. Greenlees, & C. R. Hulten (Eds.), *Price index concepts and measurement* (pp. 17–83). University of Chicago Press.

Smith, P. (2015). The importance of being timely. *Journal of Forecasting*, *34*(6), 492–506.

Smith, P. (2016). Google's MIDAS touch: predicting UK unemployment with internet search data. *Journal of Forecasting*, *35*, 263–284.

Surowiecki, J. (2011). A billion prices now. The New Yorker, May 30.

The Economist (2012). Don't lie to me, Argentina. February 25th.

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: how Google searches foreshadow housing prices and sales. In A. Goldfarb, S. M. Greenstein, & C. E. Tucker (Eds.), *Economic analysis of the digital economy* (pp. 89–118). University of Chicago Press.

Zumbrun, J. (2015). The billion prices project thinks inflation may have turned a sharp corner. *The Wall Street Journal*.

Tarek M. Harchaoui Assistant professor at the department of Global Economics and Management, University of Groningen, Netherlands, and research associate with the Groningen Growth and Development Centre and member of the NBER-CRIW.

Robert V. Janssen Business economist at a Dutch corporate finance firm.