Finding the Goldilocks data collection frequency for the Consumer Price Index[1]

Luigi Palumbo (Banca d'Italia) and Tiziana Laureti (Università degli Studi della Tuscia)

23 April 2024

Abstract

This work addresses a well-defined question: the lack of a theoretical framework to assist National Statistical Institutes (NSIs) in determining the adequacy of their data collection for the Consumer Price Index – whether it is too sparse, too frequent, or just right. We propose a framework designed to achieve a balance between reducing uncertainty in price measurement and minimizing the expenses associated with data acquisition, processing, and storage. This cost-benefit analysis is particularly relevant with the emergence of big data and alternative data sources, alongside regulatory requirements for NSIs to archive their data over extended periods. An illustrative application is provided through an examination of electricity and gas utility prices in the Italian unregulated market during year 2023, items that were notably affected by the energy price crisis stemming from Russia's invasion of Ukraine.

[1] We benefited from useful comments and inputs provided by Stefano Neri and Giordano Zevi.

1 Introduction

The collection of price data for the Consumer Price Index (CPI) should aim to strike a balance between the statistical accuracy in measuring consumer price levels and the associated costs of data collection, processing, and storage (International Labour Organization, 2020). Sparse measurements for volatile prices may lead to inaccurate estimations, while frequent measurements for stable prices could unnecessarily increase costs for the NSI responsible for CPI compilation. Our work studies the relation between price data collection frequency and volatility of month-on-month CPI changes. Combining this uncertainty with the costs for data collection, processing and storage, we propose and empirically validate a framework to determine whether a certain data collection frequency is too sparse, too frequent, or just right.

The timing of price collection is one of the sampling dimensions involved in CPI production process, which also includes product and geographic outlet[2] dimensions. For each of these dimensions, there is a universe from which a sample will be drawn. The timing of price collection is chosen purposively in most cases and the target concept is the monthly average price. The main principle is that the prices of each individual product should be collected each month at the same time, during the same week (weekdays excluding holidays) or the same day of the month. If there is some price variation within a day, it is important that prices are collected always at the same time of the day (International Labour Organization, 2020).

In Europe, NSIs are subject to Article 8 of regulation 2020/1148 for the Harmonised Index of Consumer Prices (HICP) computation, which states that observed prices shall refer to at least 1 working week at, or around, the middle of the month. In addition, if prices for an individual product are known to be volatile within a month, the observed prices shall refer to more than 1 week. Therefore, this rule establishes a minimum requirement for sampling in time as collecting all prices on a single day is not permitted and price collection must be spread over an extended period. Items like gasoline and electricity, which experience an high degree of synchronization among multiple retailers in terms of price adjustments, also necessitate data collection over an extended period.[3]

While the significance of determining the suitable frequency for data collection is evident in CPI compilation, as highlighted in various contributions (Hannon, 1998), it is surprising how scant attention the analytical determination of the optimal data collection frequency has garnered in the literature. This is particularly relevant now, as the new sources of data – such as scanner data, web scraped data, and administrative data – allow a greater frequency of price recording and broader coverage of the reporting month for the CPI compilation.

Although some stylized facts about the volatility of specific categories exist (Dhyne et al., 2006), the benefits of higher data collection frequency have been stated (Cavallo & Rigobon, 2016), and considerations regarding the impact of data collection windows on resulting indices have been explored (Haan & Opperdoes, 1997), there’s a notable absence of an analytical framework for optimizing data collection frequency. Specifically, there is no framework to help NSIs understanding if their data collection is too sparse, too, frequent, or just right. This stands in stark

[2] Geography and outlets form one dimension together since outlets are located somewhere on a map.
[3] Those requirements are extensively detailed in the updated HICP methodological manual (Eurostat, 2024).

contrast to other aspects of the CPI index production process, such as product stratification (Chessa, 2021), product sampling (Banerjee, 1956; Heravi & Morgan, 2014), or combined outlet and product sampling (Dalén & Ohlsson, 1995), where optimization frameworks abound.

This study empirically examines how the frequency of data collection influences the measurement of month-on-month CPI changes. It utilizes daily data on electricity and gas utility prices in Italy to expose the trade-off between price data collection costs and price data representativeness. The current methodology used by ISTAT, the Italian NSI, entails the centralized collection of price data in a single day every month for those utilities. The dispersion of the possible month-on-month CPI change distribution can be considered as a measure of its uncertainty, as different values may have been produced if the sampling days were to be different. Additionally, we introduce a framework for analytically determining the optimal data collection frequency for the CPI, considering the historical volatility of a particular category. Our contribution provides both theoretical and empirical assessments of the statistical uncertainty connected with CPI measurement,. This area has been largely unexplored in the literature, despite repeated calls from policy institutions for contributions (ECB, 2021).

We developed two functions to represent the cost associated with uncertainty, which decreases as the frequency of data collection increases, and the cost linked to data collection, which conversely rises with the frequency of data collection. By minimizing the combined cost functions, we can identify the optimal frequency for data collection. The precise location of the optima depends on the parameters of the cost functions, which can be adjusted to accommodate the preferences and constraints encountered by CPI practitioners in various contexts. Our objective is to enact a tool helpful for striking a balance between capturing price fluctuations and efficiently allocating resources for price data collection, processing, and storage.

2 Data

We collected price data for electricity and gas contracts daily from a government-managed web portal in Italy, where all market operators are mandated to publish their offering for new contracts on the unregulated market, from February 1, 2022 to October 31, 2022. This period was characterized by an extremely high volatility for energy prices. ISTAT uses the same web portal to acquire data for the official CPI.

For both utilities we collect prices in nine major Italian cities covering all macro-regions: Turin, Milan, Venice, Bologna, Florence, Rome, Naples, Bari, and Palermo. We use two standard consumption profiles: 2700 kWh/year for electricity and 1400 m3 for natural gas, and we collect data for both fixed and variable price contracts. For electricity we only collect prices with a flat hourly rate and for primary residences.

Operators can have multiple offers for each utility type. As a minimum, they need to offer a contract with terms and conditions (except the price) in line with the regulated energy market, called PLACET.

Despite the portal being mandatory, we observed some anomalies over the course of the data collection. Specifically, in some instance only PLACET offers were listed even if other type of contracts were available on the operator website. In order to fix those issues, which could potentially distort the index we calculate, we impute data for each operator carrying forward the last observed offer on the PLACET and non-PLACET typology in case there is no observation on the specific day and city.

We also collected data from the Italian Energy Authority on the market share of the various operators, and on the share of fixed and variable price contracts.

3 Methodology

The methodology we use in this paper to calculate the CPI is loosely based on the one used by ISTAT to calculate the official CPI for electricity and gas on the unregulated market. ISTAT collects data once during the first 15 working days each month from 20 regional capitals using differentiated consumption profiles and multiple options for the rate type (D'Amore et al., 2022). In order to calculate the official CPI ISTAT selects the major operators in each region and calculates regional price indices using the regional market shares as weights before aggregating the CPI at a national level (ISTAT, 2023). At the beginning of each quarter (January, April, July, October) ISTAT also calculates an adjustment for potential changes in the energy bonus for fragile consumers, which are modified quarterly by the Government. While this bonus used to be set annually, the Italian government implemented more frequent adjustments as response to the exceptional energy price volatility in 2022-23. Since the bonus is a rebate for specific categories of consumers, list prices do not reflect its changes.

3.1 Daily CPI calculation

We calculated daily CPI for electricity and gas using a weighted Time-Product Dummy index (Diewert, 2005; Rao, 2005), calculating first the average price for each operator each day and using the operator national market share as weight. The name TPD method was suggested by De Haan & Krsinich (2014) as it adapts Summers (1973) multilateral country-product dummy (CPD) method for spatial comparisons to price comparisons across time. The formula is:

$$log(p_{it}) = \sum_{t=1}^{T} \delta_t D_t + \sum_{i=1}^{N-1} \beta_i D_i + \varepsilon_{it}$$

Where p_{it} is the price of operator i at time t, D_t is a is a dummy which is equal to 1 for prices collected at time t and zero otherwise, D_i is a dummy equal to 1 for prices referred to operator i and zero otherwise, and $\varepsilon_i t$ is the error term. Each observation is weighted using the national market share of the operator $w_i t$. It is possible to find the CPI level at time t exponentiating the coefficient δ_t:

$$CPI_t = e^{\delta_t}$$

This calculation is performed separately for the fixed and variable price contracts, and afterward the overall CPI level for the month is calculated as weighted mean of the two sub-indexes, using the relative share of fixed and variable price contracts in the market.

3.2 Month-on-month CPI variations

Once the CPI level for each day has been calculated, we start a simulation process to find all the possible month-on-month CPI variations for different data collection frequencies, ranging from once during the first 15 business days of the month to all first 15 business days of the month.

Once the CPI level for each day has been calculated, we start a simulation process to find all the possible month-on-month CPI variations for different data collection frequencies, ranging from once during the first 15 business days of the month to all first 15 business days of the month. For each month the CPI month-on-month variation is calculated as the ration of the unweighted arithmetic mean of the CPI calculated considering the various combination of days in contiguous months:

$$\pi_t = \frac{\sum_{j=1}^{k} C\,PI_j^t}{\sum_{l=1}^{k} C\,PI_l^{t-1}}$$

Where π_t is the month-on-month CPI change, CPI_j^t are the CPI obtained from the sampled timing period of current month,, CPI_j^{t-1} are the CPI levels sampled from the previous month, and k is the number of days sampled ($k = 1, \ldots, 15$). We only calculate month-on-month changes using the same number of sampled days in contiguous months

By considering all possible combinations for each month, expressed by the binomial coefficient, Table 1 details the number of possible month-on-month variations at each sampling frequency, calculated using the second power of the binomial coefficient:

$$C(15, k)^2 = \binom{15}{k}^2 = \left(\frac{15!}{k!\,(15-k)!}\right)^2$$

Where k is the number of sampled days during the first 15 business days each month. It is possible to see in Table 1 how this number of combinations grows very large for combinations of 7 and 8 days. Considering both utilities, this study entails the calculation of over 2.5 billion possible month-on-month CPI changes, with a considerable use of IT resources. For all months and all frequencies, after calculating all the possible CPI changes, we calculate the standard deviation and coefficient of variation of the possible month-on-month variations.

Table 1: Number of possible month-on-month CPI variation for each data collection frequency.

Number of sampled days	Possible CPI monthly levels	Possible CPI month-on-month variations
1	15	225
2	105	11,025
3	455	207,025
4	1,365	1,863,225
5	3,003	9,018,009
6	5,005	25,050,025
7	6,435	41,409,225
8	6,435	41,409,225
9	5,005	25,050,025
10	3,003	9,018,009
11	1,365	1,863,225
12	455	207,025
13	105	11,025
14	15	225
15	1	1

3.3 Optimization framework

In order to strike a balance between the accuracy of price level measurement and the efficiency of data collection, processing, and storage, several key considerations merit attention. Firstly, it is crucial to establish an appropriate relationship between the frequency of data collection and the uncertainty surrounding the month-on-month changes in the CPI. This can be achieved by utilizing empirical standard deviations or coefficients of variation derived from all the possible CPI month-on-month changes, calculated with the methodology outlined in Section 3.2 to determine the most suitable functional relationship expressing measurement uncertainty and data collection frequency. To compare different functional forms of cost functions, a regression analysis is conducted using a variety of linear and non-linear functions, with the $adjusted - R^2$ serving as the selection criterion.

The second aspect involves aligning both the uncertainty and the cost associated with measurement using a common metric. Two potential approaches emerge. One approach involves assigning a monetary value to each unit of measurement uncertainty, recognizing its adverse impact on NSIs, which may then allocate financial resources to mitigate it. Alternatively, the cost of data collection could be converted into an equivalent uncertainty value, representing the minimum reduction in uncertainty sought by the NSI through additional data collection efforts. Both approaches are viable and, if consistently parameterized, should yield identical outcomes. For clarity, we adopt a predetermined uncertainty value for each instance of data collection, symbolizing the minimum reduction in uncertainty desired by the NSI through supplementary data collection activities.

Thirdly, it is necessary to establish a window for data collection. Consistent with the current methodology of ISTAT, we designate the first 15 business days of the month as the potential period for data collection. Several reasons support this choice. Firstly, no systematic differences in prices between business days and weekends were observed (a feature typical of these items). Secondly, conducting data acquisition during weekends would either increase costs due to additional personnel shifts or raise the risk of missing data if the acquisition process were left unmonitored, without providing substantial benefits.

Once the three aspects above have been cleared, we can formalize a discrete minimization problem for the overall cost function given the number of data collection occurrences in the month as follow:

$$min\big(UncertaintyCost(k) + CollectionCost(k)\big), k \in \mathbb{N}^{+}: \{1 \leq k \leq 15\}$$

The optimal number of data collection occurrences k is an integer between 1 and 15, included, which minimizes the combined cost of uncertainty and data collection.

4 Results

Figure 1 illustrates the outcomes of our daily CPI calculations for electricity and gas using February 1, 2024 – the first day in our series – as reference level. For comparison we include the official price indices for both utilities, rescaled using February 2024 as reference level. While there are noticeable differences between our indices and the official ones, the overarching trends exhibit considerable similarity. The largest difference is observed in April, when the energy bonus for fragile consumers was considerably revised. Since our daily index does not consider the effect of the bonus, the divergence is easily explained. Notably, our daily indices demonstrate remarkable volatility, particularly evident in the context of gas prices. Specifically, there is a substantial number of instances where consecutive-day differences surpass 5 percentage points.

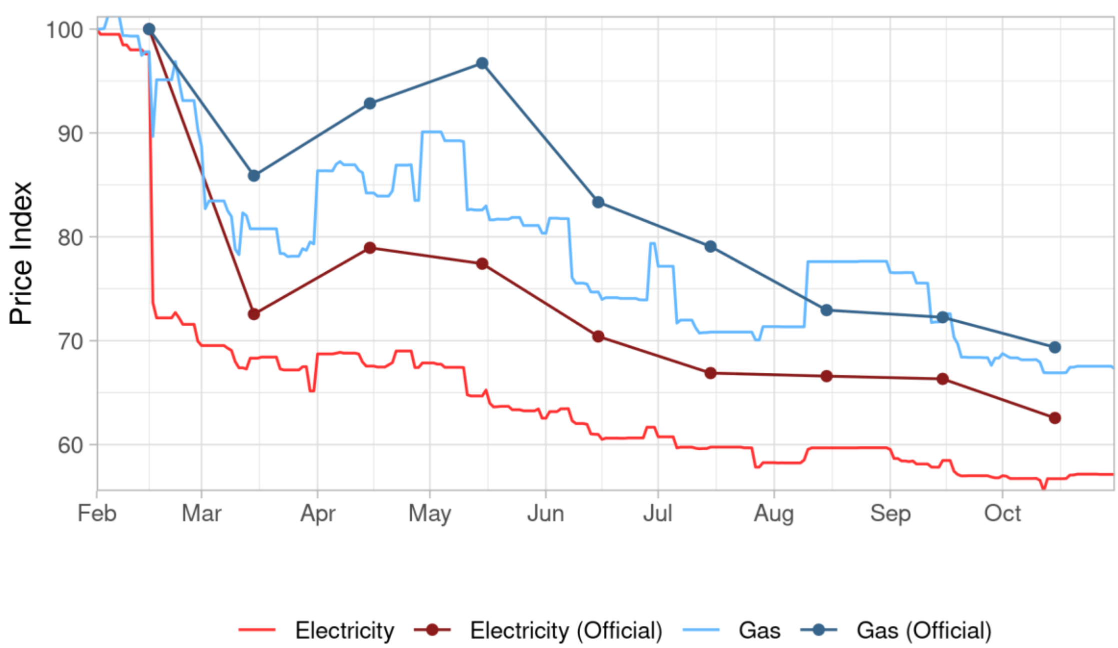

Figure 1: Price indices for Electricity and Gas. Daily price indices rescaled with 2024-02-01=100. Official price indices rescaled with 2024-02=100. Official price indices represented in the middle of the month to which they refer.

Given this high volatility, it would not be a surprise that month-on-month CPI variations with a single data collection day each month may be very different according to the effective data collection day. In fact, in many cases they may well be of different sign. Figure 2 show the results on March 2023 for electricity and gas, respectively. As expected, we can note that the dispersion of potential month-on-month CPI variations decreases as the data collection frequency increases, until reaching zero dispersion when data is collected every day, covering the target universe in time dimension. The median value seems to be stable for data collection frequencies higher than 4 times per month but presents substantial swings at lower frequencies. Figures 6 and 7 in Appendix A shows the same results for each month in our sample. In many cases we can note an interquartile range larger than 2 percentage points.

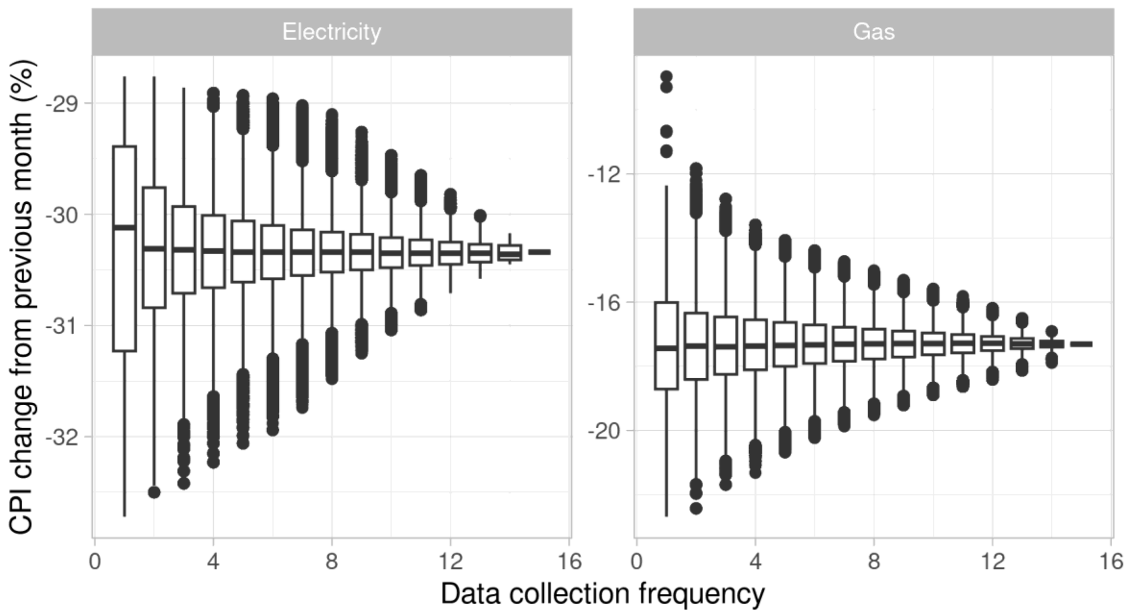

Figure 2: Dispersion of potential month-on-month CPI change compared to previous month according to the data collection frequency, March 2023.

Figure 3 shows the density distribution of the potential month-on-month CPI changes in March 2023 at selected frequencies for electricity and gas, respectively. Figures 8 and 9 in Appendix A shows the same results for each month in our sample. We observe that also at high frequencies there may be more than one mode. However, the reduction of overall dispersion with each additional data collection occurrence is noticeable.

While the graphics in Figures 6 to 9 provide us with insights on the overall trend and relationship between data collection frequency and uncertainty for month-on-month CPI changes, we need a synthetic metric to perform further analysis. Figures 10 and 11 in Appendix A provide us a comparison of the two metrics we tested for this task for electricity and gas: standard deviation and coefficient of variation. While both metrics exhibit similar trends, the patterns for the coefficient of variation seem to have considerably higher dispersion at lower data collection frequencies. Given the information presented in Figures 6 and 7, it seems that the average value used to calculate the coefficient of variation may effectively impact the pattern of the metric over different sampling frequencies. Therefore, we selected the standard deviation as metric for our modeling exercise. It should also be noted that the CPI standard deviation distribution for gas shows higher values, remarking the considerably higher volatility in gas prices over the period studied.

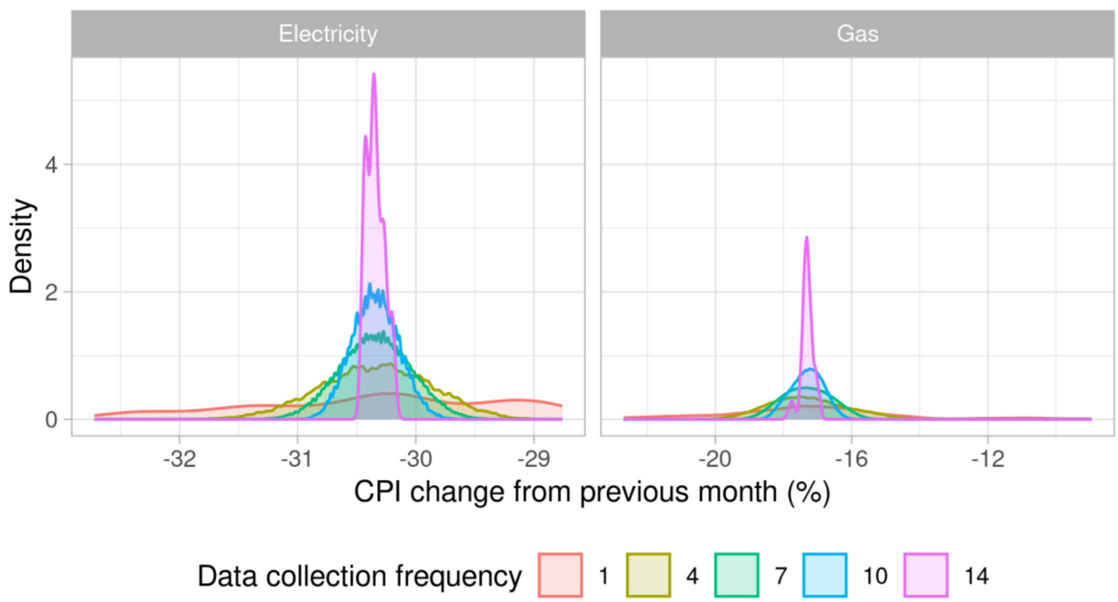

Figure 3: Density distribution of potential month-on-month CPI change compared to previous month according to the data collection frequency, March 2023.

Considering the trends in Figure 10, we tested the following specification for our uncertainty cost function, as the data seems to approximately follow an hyperbolic distribution:

$$UncertaintyCost(k) = \beta_0 + \beta_1 k + \frac{\beta_2}{k} + \eta$$

Where k is the data collection frequency per month, and η is the error term. For robustness, we conducted additional tests excluding alternate forms that did not involve both $\beta_1$and β_2 simultaneously. Results are summarized in Table 2.

Table 2: Uncertainty cost function: regressions results. Preferred specification shaded in gray.

	Standard Deviation					
	Electricity			Gas		
	(1)	(2)	(3)	(4)	(5)	(6)
β_0	0.147***	1.167***	0.614***	0.394***	3.117***	1.649***
	(0.029)	*(0.049)*	*(0.075)*	*(0.073)*	*(0.124)*	*(0.182)*
β_1		-0.084***	-0.042***		-0.224***	-0.114***
		(0.005)	*(0.006)*		*(0.014)*	*(0.016)*
β_2	1.581***		1.002***	4.218***		2.664***
	(0.090)		*(0.117)*	*(0.225)*		*(0.283)*
Adjusted-R2	0.720	0.668	0.794	0.746	0.694	0.824
Note:	*p<0.1; **p<0.05; ***p<0.01					

For all the alternative specifications we tested coefficients are significant, consistent with our expectations, and coherent between the two utilities. Since we decided to use the adjusted-R^2 as selection criteria, we select the full specification as cost function for the prosecution of our analysis. We note that the ordering of the different specifications is consistent across electricity and gas. In Figure 3 we plotted the model results against the standard deviation historical values for both utilities. Overall, we consider this cost function a good approximation of our data.

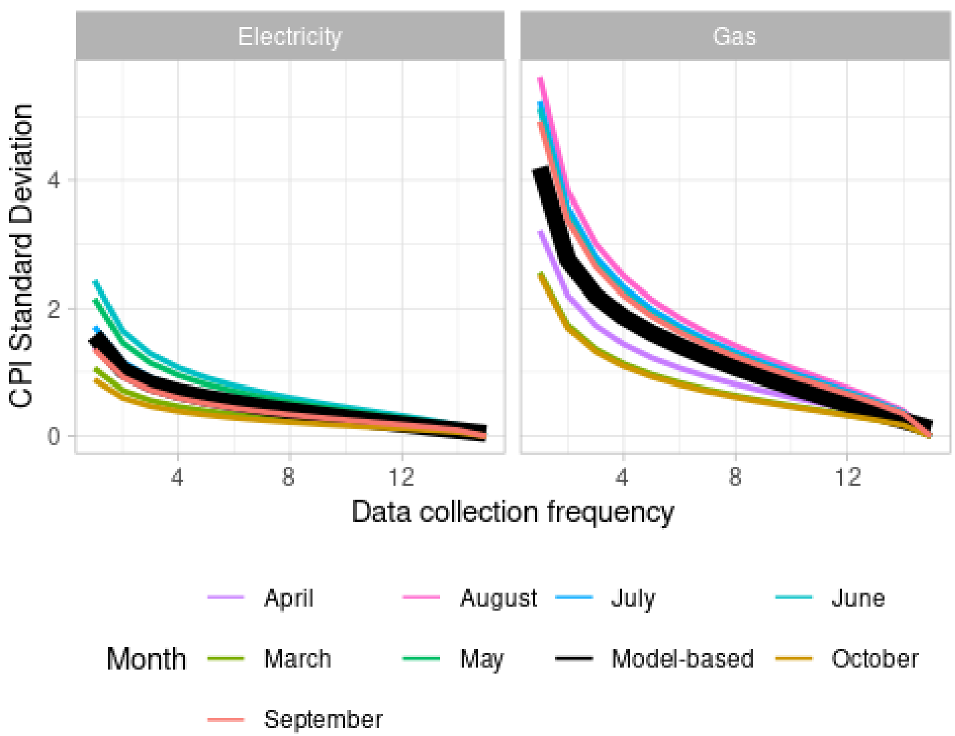

Figure 4: Electricity and gas: Standard Deviation for month-on-month CPI changes at different data collection frequencies. Black tick line shows the model-based month-on-month CPI uncertainty parametrized for each utility.

The concluding step in our proposed optimization framework entails the allocation of a cost to each instance of data collection, reflecting the minimum anticipated reduction in uncertainty upon its implementation. With our selected metric of uncertainty being the standard deviation of month-on-month CPI variations, presented as a percentage, this cost will be represented in the identical unit of measurement. To provide clarity, we examine a range of cost values expressed as percentage points per occurrence of data collection for electricity and gas, serving as illustrative examples. Therefore, our data collection cost can be expressed as follow:

$$CollectionCost(k) = \gamma k, k \in \{0.1,0.2,0.3\}$$

Where k is the data collection frequency per month and γ is the cost per data collection occurrence. However, it is important to note that this definition is neither universally applicable nor entirely objective. Each NSI should conduct its own assessments, taking into account the actual expenses incurred for data acquisition, processing, and storage, as well as their own tolerance to uncertainty.

Based on the selected parameterization grid, we computed the costs associated with uncertainty and data collection, as well as their aggregate value. Figure 5 shows the results of our analysis, with the different optimal points for electricity and gas according to different γ parameters.

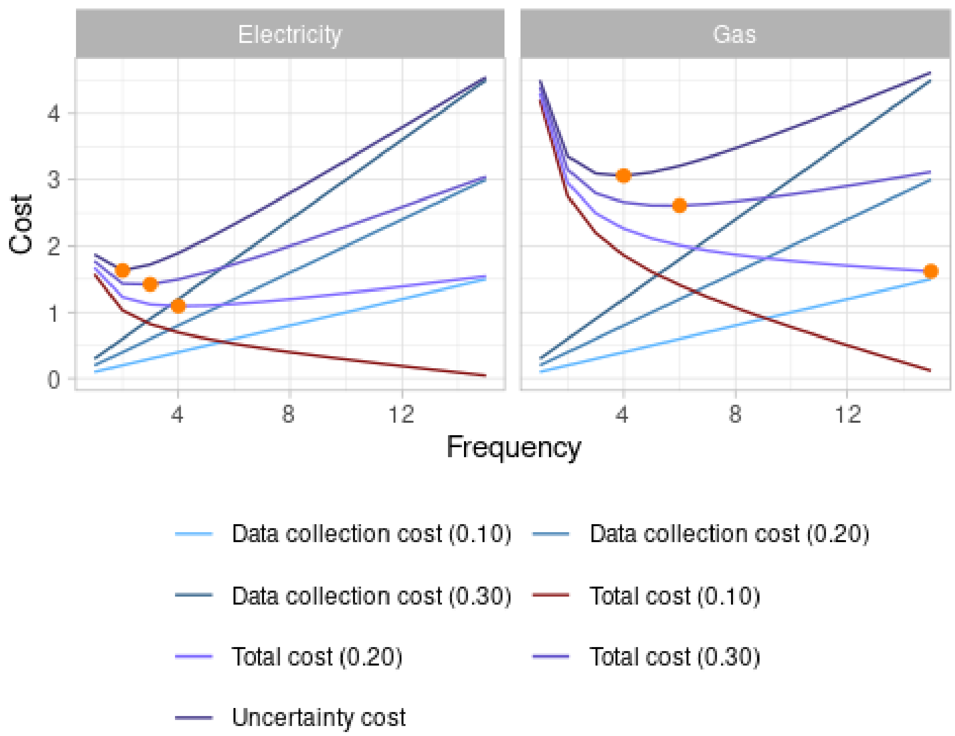

Figure 5: Electricity and gas: Uncertainty cost, data collection cost, and total cost. Orange dots mark the minimum total cost with the selected parametrization for each utility.

The optimal data collection frequency for gas, which prices exhibited a larger volatility during the period studied, results higher than the one for electricity at every value of γ. This result is consistent with the general recommendation to perform more frequent data collection on products and services that exhibit higher price volatility. While the specific results are highly dependent on the selected parametrization, given the shapes of the cost functions we are confident that a minimization procedure would be generally viable.

Several caveats should be considered when assessing our findings. First of all, using historical volatility as a proxy of the future one is a good guidance only as far as there are no disruptions in the price patterns. The volatility for energy prices in Europe during 2022-23 was extremely large and not anticipated. NSIs should routinely reassess the volatility of a given category to ensure the data collection frequency stays optimal. Secondly, our empirical results for electricity and gas prices should not be compared directly with the official data published by the Italian NSI. We use a different – much smaller – coverage across space and rate options, a single national consumption profile rather than customized for each region, and a completely different methodology for index calculation. Finally, the translation of uncertainty and data collection costs in a common unit of measure is far from being an objective operation. Our exercise is exemplificative, and each NSI should judiciously study the trade-off between those dimensions according to its resources and objectives.

5 Conclusions

The paper derives analytically the optimal data collection frequency for CPI on a given category, using its past volatility as guidance. It also proves empirically that a sampling frequency much lower than optimal may effectively create large uncertainty in the month-on-month CPI variations. On the other hand, increases in the CPI data collection frequency yield diminishing returns. It is important for a NSI to understand at which point is no longer cost-efficient to perform additional data collection exercises.

We believe that data collection frequency is a key driver for the quality of CPI compilation, and it deserves careful consideration while designing the CPI methodology. The vertiginous expansion of high frequency data sources, such as scanner data, web scraping, and APIs offers great opportunities for CPI practitioners and academics. Today, even with limited resources and only leveraging publicly available data, it is possible to perform targeted studies aimed to assess the fitness of current CPI methodologies and propose improvement based on empirical evidence. Those opportunities are paired with substantial IT challenges and costs connected to the management of huge volume of data. In conclusion, we believe that empirical validation of CPI methodology with the use of high frequency data sources is an area with high potential for further studies, and we are looking forward to more contributions in the near future.

References

Banerjee, K. S. (1956). A note on the optimal allocation of consumption items in the construction of a cost of living index. *Econometrica*, *24*(3), 294–295.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, *30*(2), 151–178. https://doi.org/10.1257/jep.30.2.151

Chessa, A. G. (2021). A product match adjusted r squared method for defining products with transaction data. *Journal of Official Statistics*, *37*(2), 411–432. https://doi.org/10.2478/jos-2021-0018

D'Amore, A., Duma, R., Dormi, R., C., & Simone, A. (2022). Web scraping techniques and API to estimate italian inflation. *Room Papers, Seventeenth Meeting of the Ottawa Group*.

Dalén, J., & Ohlsson, E. (1995). Variance estimation in the swedish consumer price index. *Journal of Business & Economic Statistics*, *13*(3), 347–356. https://doi.org/10.1080/07350015.1995.10524608

De Haan, J., & Krsinich, F. (2014). Scanner data and the treatment of quality change in nonrevisable price indexes. *Journal of Business & Economic Statistics*, *32*(3), 341–358. https://doi.org/10.1080/07350015.2014.880059

Dhyne, E., Alvarez, L. J., Le Bihan, H., Veronese, G., Dias, D., Hoffmann, J., Jonker, N., Lunnemann, P., Rumler, F., & Vilmunen, J. (2006). Price changes in the euro area and the united states: Some facts from individual consumer price data. *Journal of Economic Perspectives*, *20*(2), 171–192. https://doi.org/10.1257/jep.20.2.171

Diewert, E. (2005). Weighted country product dummy variable regressions and index number formulae. *Review of Income and Wealth*, *51*(4), 561–570. https://doi.org/10.1111/j.1475-4991.2005.00168.x

ECB. (2021). Inflation measurement and its assessment in the ECB's monetary policy strategy review. *ECB Occasional Paper Series, 265.*

Eurostat. (2024). *Harmonised index of consumer prices (HICP): Methodological manual*. Publications Office of the European Union. https://doi.org/10.2785/055028

Haan, J. de, & Opperdoes, E. (1997). Estimation of the coffee price index using scanner data: The choice of the micro index. *Third Meeting of the Ottawa Group*, 16–18.

Hannon, J. B. (1998). Data collection and quality assurance problems in consumer price indices. *Statistical Journal of the United Nations Economic Commission for Europe*, *15*(1), 85–96. https://doi.org/10.3233/sju-1998-15105

Heravi, S., & Morgan, P. (2014). A comparison of six sampling schemes for price index construction in a COICOP food group. *Applied Economics*, *46*(22), 2685–2699. https://doi.org/10.1080/00036846.2014.909574

International Labour Organization. (2020). *Consumer price index manual, 2020: Concepts and methods*. ILO Publications. https://doi.org/10.5089/9781484354841.069

ISTAT. (2023). *Gli indici dei prezzi al consumo*. ISTAT.

Rao, D. P. (2005). On the equivalence of weighted country-product-dummy (CPD) method and the rao-system for multilateral price comparisons. *Review of Income and Wealth*, *51*(4), 571–580. https://doi.org/10.1111/j.1475-4991.2005.00169.x

Summers, R. (1973). International price comparisons based upon incomplete data. *Review of Income and Wealth*, *19*(1), 1–16. https://doi.org/10.1111/j.1475-4991.1973.tb00870.x

Appendix A

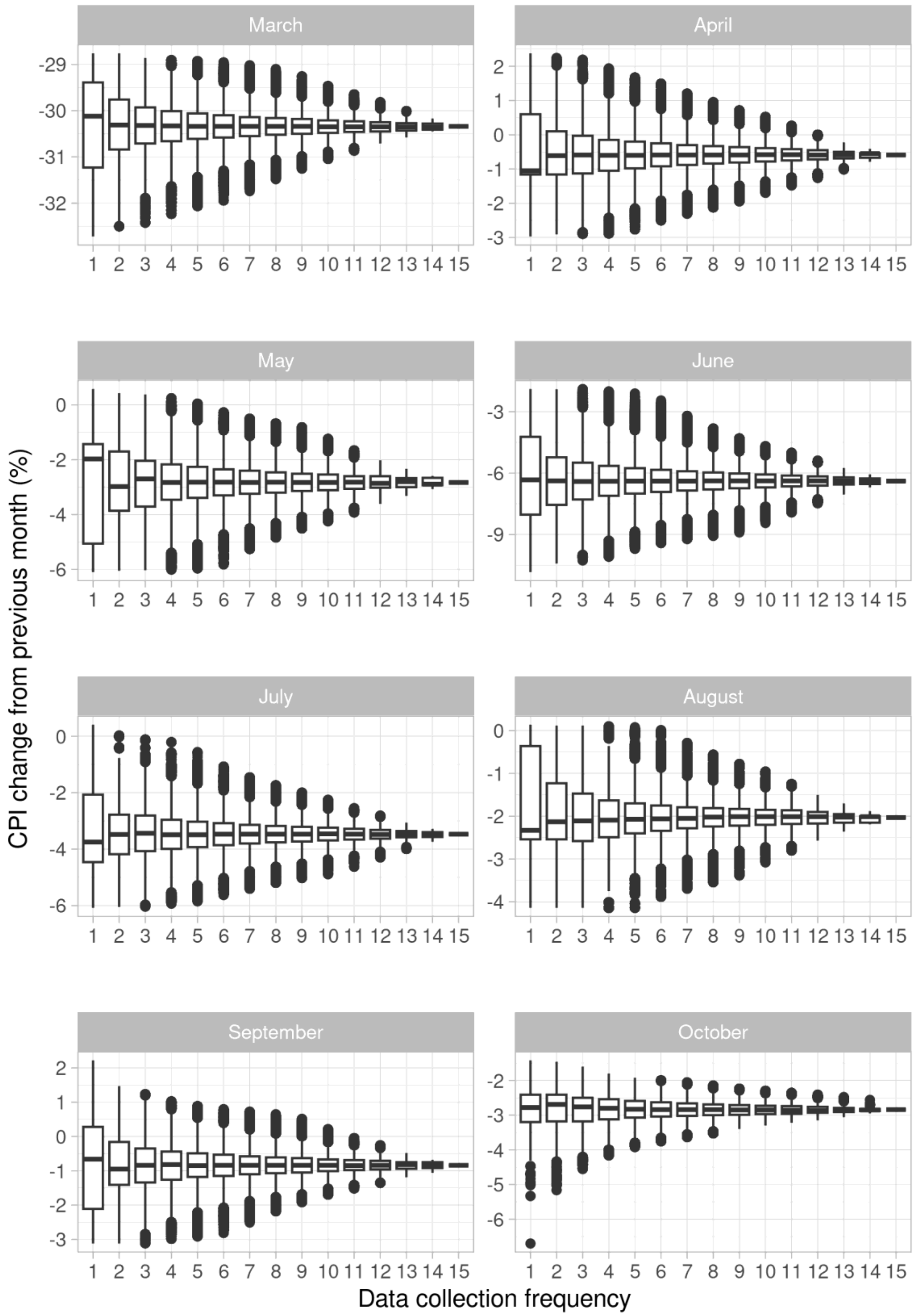

Figure 6: Electricity - Dispersion of potential month-on-month CPI change compared to previous month according to the data collection frequency.

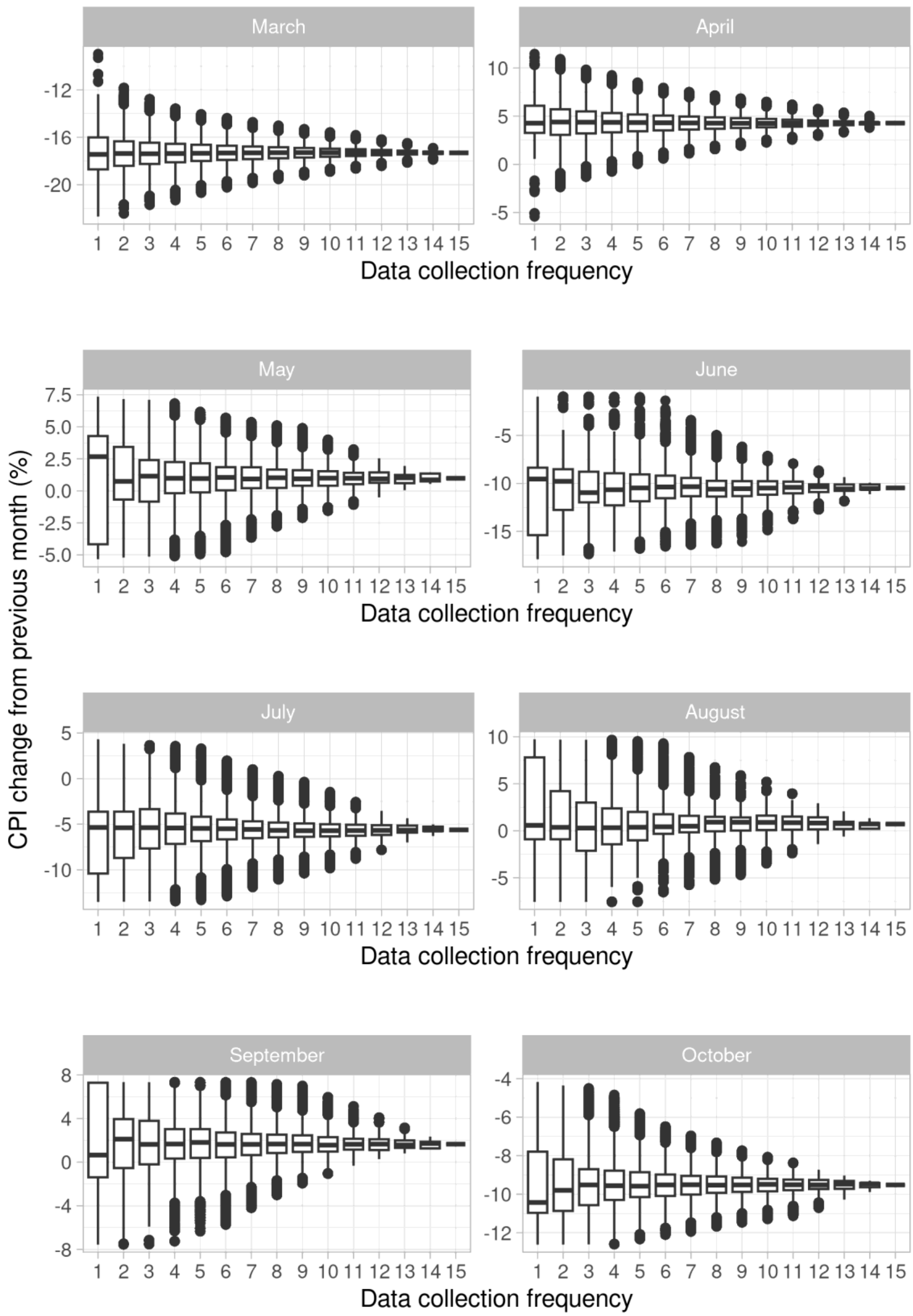

Figure 7: Gas - Dispersion of potential month-on-month CPI change compared to previous month according to the data collection frequency.

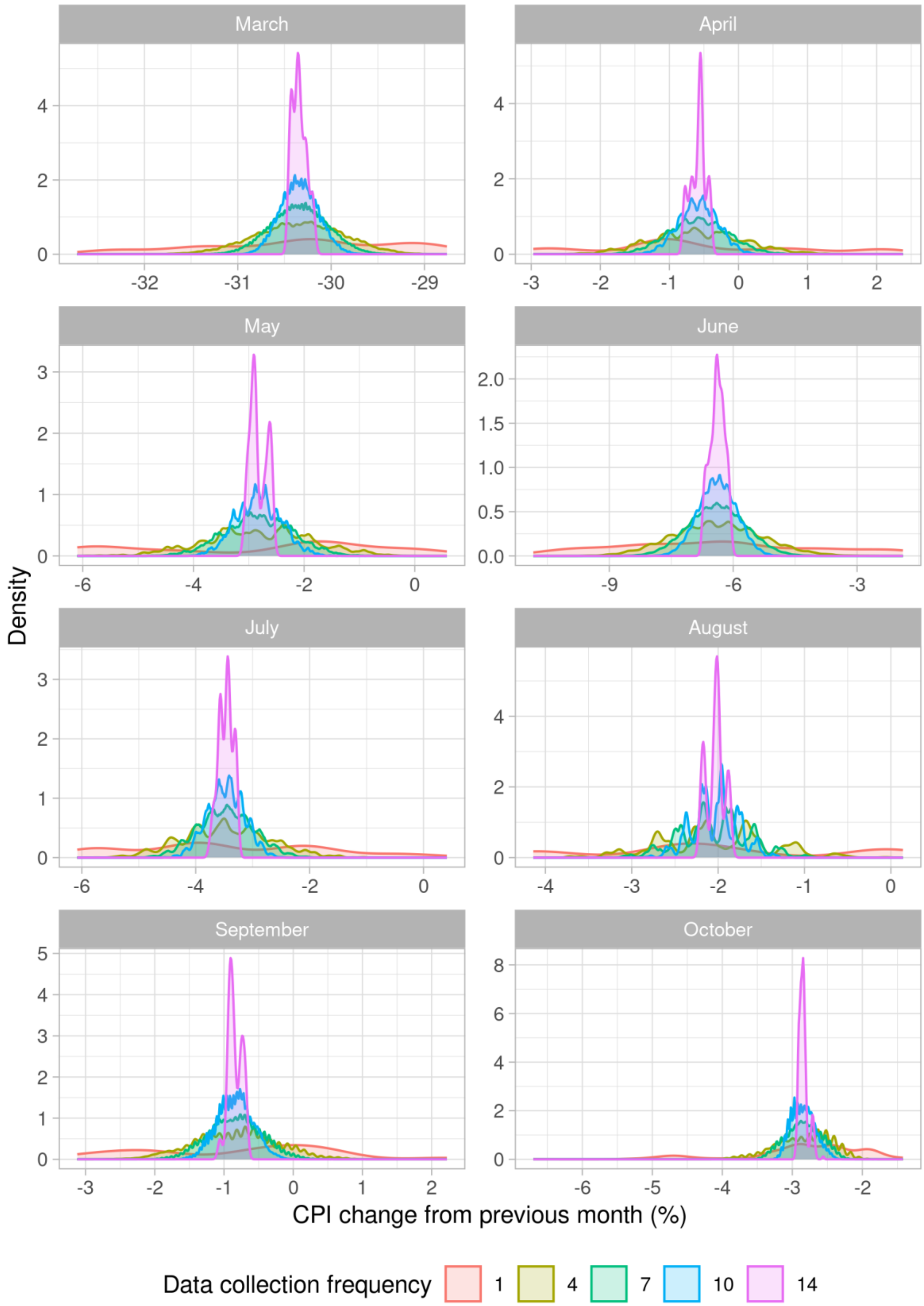

Figure 8: Electricity - Density distribution of potential month-on-month CPI change compared to previous month according to the data collection frequency.

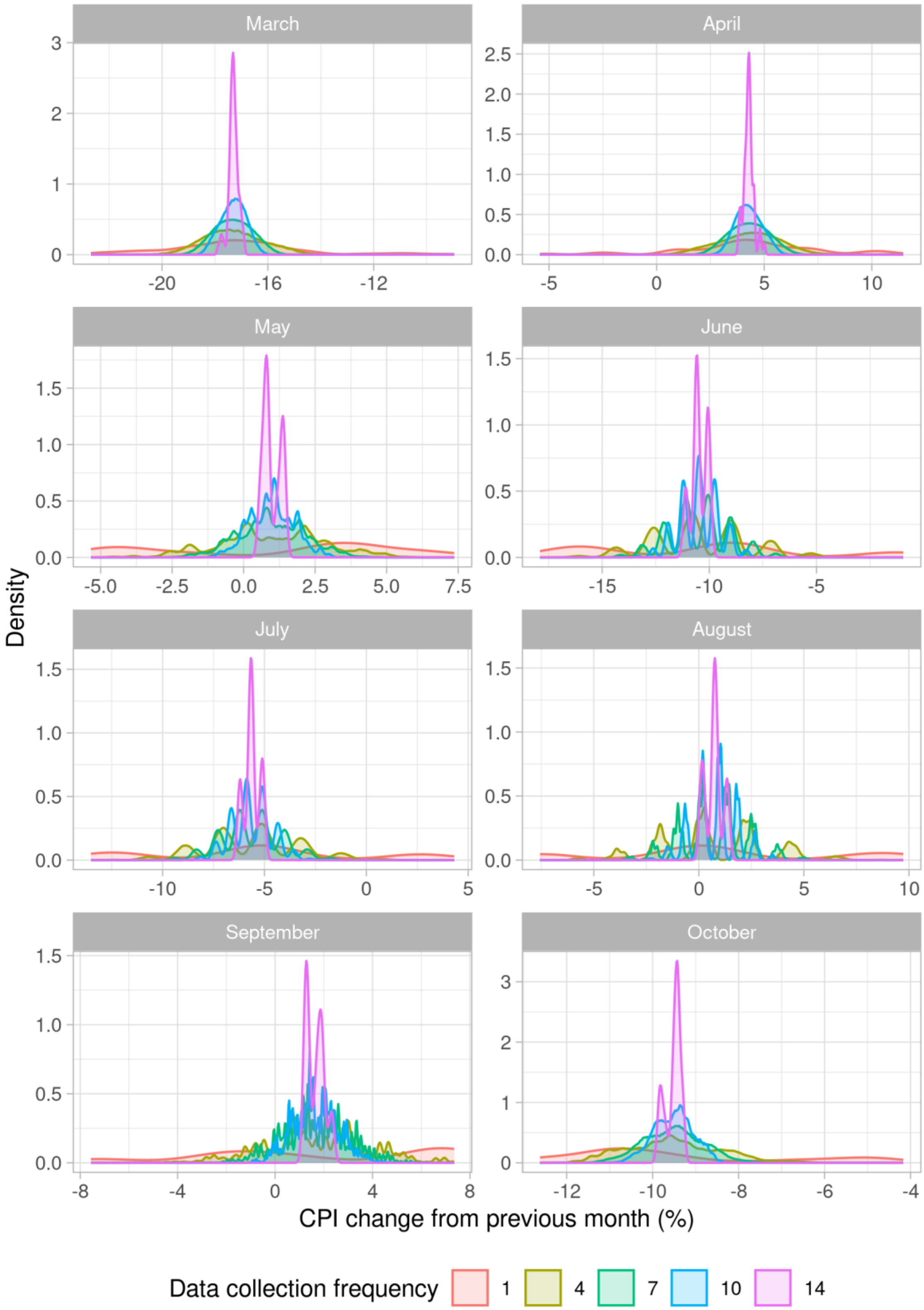

Figure 9: Gas - Density distribution of potential month-on-month CPI change compared to previous month according to the data collection frequency.

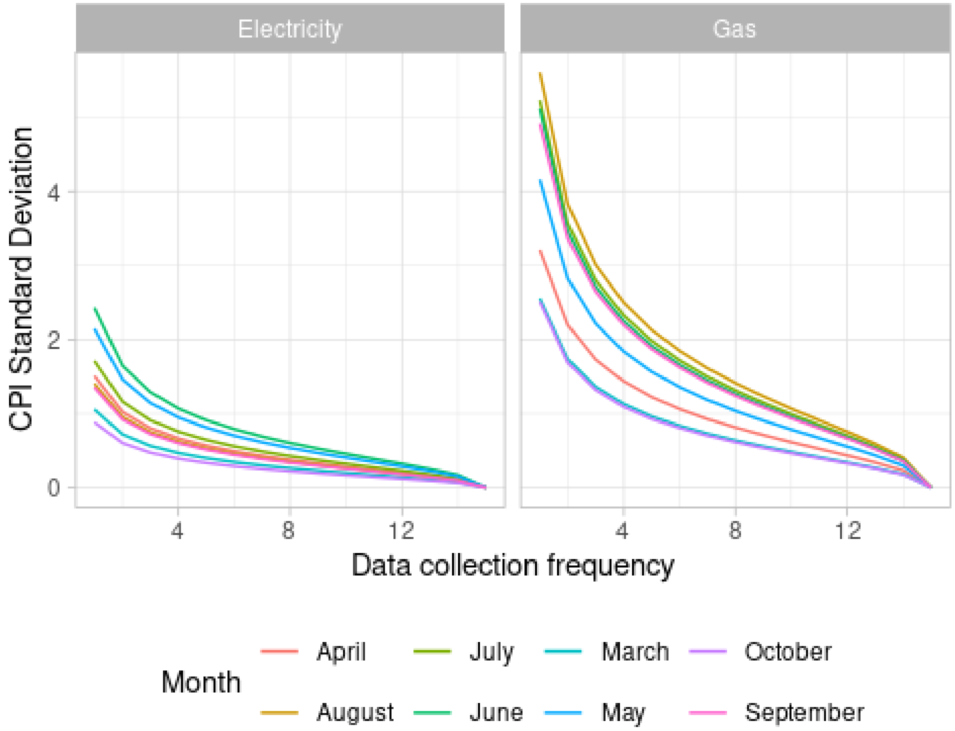

Figure 10: Electricity and gas: Standard Deviation for month-on-month CPI changes at different data collection frequencies.

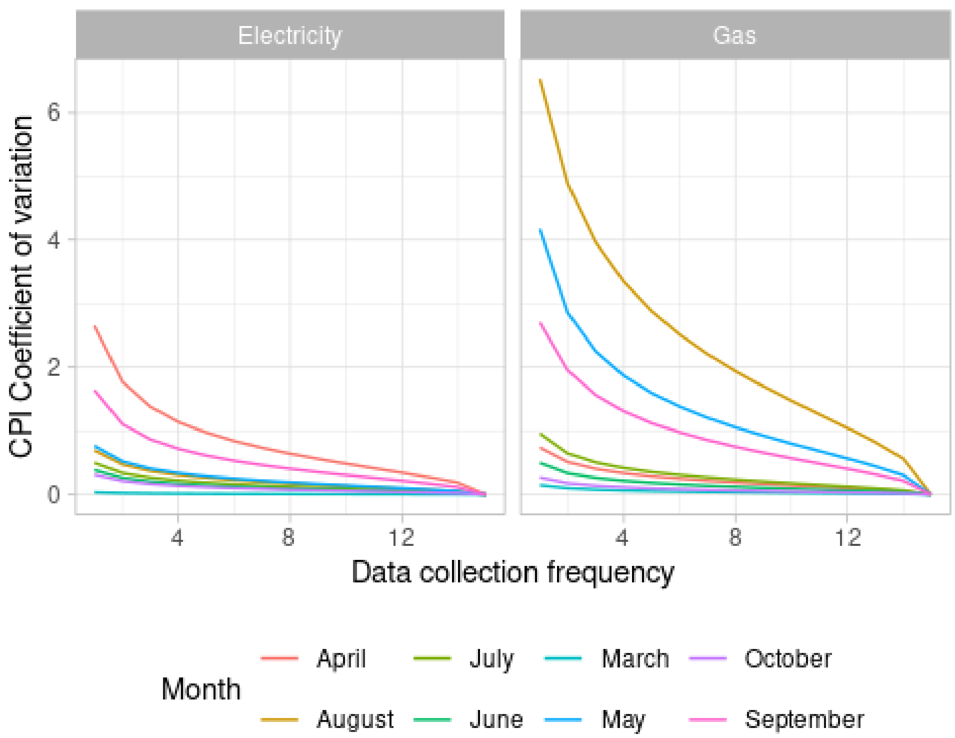

Figure 11: Electricity and gas: Coefficient of variation for month-on-month CPI changes at different data collection frequencies.