

# A NEW APPROACH FOR INTEGRATING TRANSACTION DATA IN THE CELLULAR SERVICES PRICE INDEX (CSPI)

STATISTICS CANADA

## TABLE OF CONTENTS

Overview .....	2
Scope of Cellular services .....	2
Outlets .....	2
Data .....	3
Web collected data .....	3
Transaction data .....	3
Data processing .....	4
Web collected data .....	4
Transaction data .....	4
Quality adjustment for new plans and service packages .....	7
Web collected data .....	8
Transaction data .....	8
Aggregation .....	9
Web collected data .....	9
Transaction data .....	9
Hybrid indexes .....	9
Results .....	10
Integration timelines and plans for operationalization .....	10
Conclusion .....	11
Appendix .....	12
Application of Machine Learning methods for hedonic quality adjustment .....	12
Application of the MARS model .....	13

## OVERVIEW

The Cellular services price index (CSPI) is one of two components of the Telephone services sub index in the Consumer Price Index (CPI), and it represents **1.33%** of the 2022 CPI basket.<sup>1</sup>

Statistics Canada continues to prioritize data accuracy, quality, and timeliness in measuring price change and producing a CPI that reflects the experience of Canadians. Statistics Canada has been working with the major Canadian wireless services providers (WSPs) to obtain transaction data for wireless plans. This initiative is aimed at improving the quality of the CSPI by expanding the sample and incorporating hedonic quality adjustment methods and plan level weights.

Thanks to this collaboration, Statistics Canada now receives monthly transaction data containing plan level information from wireless service providers (WSPs) with several retail brands across Canada. Given that not all WSPs in the market have agreed to provide the agency with transaction data, the CSPI will continue to be built on transaction data and web collected data.

This document details the proposed methodology that will be used to incorporate the new transaction data in the CSPI. The result is a “hybrid” index that combines the transaction data with web collected data from the remaining WSPs in the sample<sup>2</sup>.

## SCOPE OF CELLULAR SERVICES

For the CSPI, the product being priced includes all costs associated with the monthly services of a cellular device. These services include local and long-distance voice calls, text messaging, and internet access on a cellular device. It does not include the cost of the cellular device. Other miscellaneous fees, such as activation fees, overage charges, and roaming charges are also excluded.<sup>3</sup> It is important to note that over the years, the services included in the cellular services component of the CPI have evolved with the consumption patterns of Canadians.

## OUTLETS

In each province or territory, the largest WSPs by revenue share are sampled. Within each province or territory, WSPs are weighted by their provincial or territorial market share, calculated in terms of revenue. Weighting the WSPs by their market share aims at improving the representativeness of the index. The revenue is obtained from the [Annual Survey of Telecommunications](#) (AST), conducted jointly by Statistics Canada and Canadian Radio-Television and Telecommunications Commission (CRTC). The WSPs’ weights at provincial and territorial level are regularly updated to reflect their current market shares.

---

<sup>1</sup> The two components of Telephone services are Local and long-distance telephone service (not published), and Cellular services.

<sup>2</sup> The methodology for the current Cellular services price index is described here: [New approach for estimating the Telephone Services Index of the Consumer Price Index \(statcan.gc.ca\)](#).

<sup>3</sup> These miscellaneous fees either make up a small portion of total mobile retail revenues or there isn’t sufficient information to weigh them (as is the case with activation fees).

## WEB COLLECTED DATA

Canadian consumers have few WSPs to select from, and each WSP offers a limited number of cellular service plans on its website. Given the dynamic nature of the market, the consumer profile method is used, where prices for several consumer profiles, or representative bundles of cellular services, are tracked over time.<sup>4</sup> The method of pricing the same consumer profile over time ensures constant quantity and quality of services, and that the index reflects pure price change. The set of consumer profiles in use are designed to best reflect cellular services usage patterns of Canadian households and are regularly updated to ensure that they remain relevant. New profiles can be added based on market share shift or technological change in the telecommunication industry. Table 1 includes two examples of representative consumer profiles.

**Table 1. Fictional examples of two representative consumer profiles**

Service	Profile A	Profile B
<b>Nationwide calling minutes</b>	Unlimited	Unlimited
<b>Nationwide text messaging (SMS)</b>	Unlimited	Unlimited
<b>Data (GB)</b>	[1, 15)	[50, 100)
<b>Network type</b>	5G	5G+
<b>Canada-US calling</b>	No	Yes

Each consumer profile is matched with one plan offered by a WSP in each province. Plan prices are collected each month from the websites of WSPs. Prepaid plans are excluded from collection due to their low market share.<sup>5</sup>

## TRANSACTION DATA

Each month, Statistics Canada receives a file from participating WSPs that contains information on wireless plans from each of their retail brands across Canada. All in-market postpaid and prepaid wireless plans, and the top 100 postpaid and top 10 prepaid wireless plans (by subscriber count) are included.<sup>6</sup> Collectively, the coverage represents 70% of the WSP's subscriber base.

For each plan, the following features are provided:

<sup>4</sup> This approach is in line with international standards as per the [Consumer Price Index Manual: Concepts and Methods](#), published by the International Monetary Fund. The consumer profile method is one of the most used methods to measure price change for telecommunications services by [Member States of Eurostat](#).

<sup>5</sup> As per the 2022 Annual Survey of Telecommunications, 94% of retail mobile revenues is generated from postpaid plans.

<sup>6</sup> An in-market plan is a plan that is available for purchase by consumers, while a legacy plan is a plan that has been phased out of the market and is no longer offered to subscribers. Note that an in-market plan is not limited to only the plans that are advertised on the WSP's website and can include unadvertised plans, such as "winback" plans offered to former subscribers who have moved their services to a different WSP. A prepaid plan, often referred to as a "pay-as-you-go" plan, lets the consumer purchase a finite amount of calls/messages/data upfront, whereas a postpaid plan lets the consumer pay for a bundled set of services (i.e., calls/messages/data) after usage.

**Table 2. Feature names and descriptions of plans included in the transaction data**

	Feature name <sup>7</sup>	Description
1.	Ranking number	a plan’s rank based on its subscriber count in a province or territory
2.	Geography	the province or territory where a plan is offered
3.	Plan ID	an internal WSP-generated alphanumeric code used to identify plans
4.	Plan Name	the name of the plan—will sometimes include a description of the plan’s characteristics (like number of minutes, etc.)
5.	Retail Brand	the name of the retail brand offering the cellular service plan
6.	Plan Type	indicator for prepaid or postpaid plans
7.	Plan Availability	indicator for in-market or legacy plans
8.	Monthly Recurring Charge (MRC)	monthly price of a plan, before tax and other fees
9.	Voice	voice allowance, in number of minutes per month
10.	Short Message Service (SMS)	SMS allowance, in number of texts per month
11.	Data	data allowance, expressed in Gigabytes (GB) per month
12.	Subscriber count	count of subscribers on the plan

All plan features (except for Plan Availability and Subscriber count) for a given plan remain constant over time. A plan’s availability status could change from ‘in-market’ to ‘legacy’ when the retail brand stops advertising it on its website or offering it to new subscribers. Any modification by the retail brand to an existing plan is introduced in the dataset as a new observation with a new Plan ID.<sup>8</sup>

## DATA PROCESSING

### WEB COLLECTED DATA

Once a plan has been selected for a consumer profile, it will be priced every month until the WSP either stops advertising it or makes significant modifications to its features (e.g., adding bonus data). In both situations, a quality adjustment is made before proceeding.

### TRANSACTION DATA

#### TRACKING PLANS TO MANAGE CHURN

To track plans over time, a “unique plan ID” is created for each observation in the dataset by concatenating the Plan ID, Geography, Retail Brand, and Plan Type features.<sup>9</sup> Note that the ‘Plan Availability’ feature is not included in this derived unique plan ID, because WSPs routinely stop offering certain plans to new subscribers but will continue to provide services to active subscribers on these “legacy” plans. When this happens, the value of the

<sup>7</sup> The agency is interacting with WSPs to obtain data on new price determining characteristics (such as network type, upload/download speeds, etc.) as the market changes.

<sup>8</sup> For example, if a Retailer offers a promo on its website for an \$80 10 GB plan, reducing its price to \$70, the cheaper promotional plan will appear as a separate plan or row in the dataset. We can reflect this price decline in the index by assigning the two separate plans in the dataset as a single observation, referred to as a service package. See the [Assigning comparable plans to service packages](#) section.

<sup>9</sup> The Plan ID is not unique for every plan (or row) in the dataset, which is why this step is necessary.

Plan Availability feature switches from “in-market” to “legacy”. Including this feature in the unique plan ID feature would impede tracking of plans over time.

Given that the inclusion of any legacy plan in the file is contingent on its popularity, legacy plans that fail to reach the top 100 (for postpaid) or top 10 (for prepaid) threshold will naturally drop out of the monthly files. At the same time, popular in-market plans that become legacy plans in the current month may appear in the list of the top 100 legacy postpaid plans in the following month. This dynamic over time results in plan-level churn. It is important to note that the disappearance of a legacy plan from the monthly file does not mean that the plan no longer has any subscribers. The top 101 postpaid legacy plan (or top 11 prepaid legacy plan) will still have active subscribers, although it will not appear in the monthly files.

To reduce the impact of churn on the index, the following actions are taken during data processing, grounded on information provided by WSPs that plan features do not change once a plan is launched in the market and has active subscribers.

- 1) **Carry forward:**
  - a. **Legacy plan:** A legacy plan that is in the previous month ( $T_0$ ) but not in the current month ( $T_1$ ) gets its feature information (e.g., data allowance, MRC, etc.) carried forward to  $T_1$ . The plan’s disappearance in  $T_1$  is likely due to the plan losing subscribers between the two periods. For example, the missing legacy plan is postpaid, its  $T_0$  subscriber counts are carried forward to  $T_1$  if they are less than the subscriber counts of the top 100 postpaid legacy plan in  $T_1$ . Otherwise, the subscriber counts of the top 100 postpaid legacy plan in  $T_1$  is used.<sup>10</sup>
  - b. **In-market plan:** In-market plans missing in  $T_1$  but present in  $T_0$  have their feature information carried forward to  $T_1$ . These missing plans likely had their Availability switched to “legacy”, and thus, continue to exist in the market but were not popular enough to be included in the top 100 sample of postpaid legacy plans (or top 10 sample of prepaid legacy plans) in  $T_1$ .
- 2) **Carry backward:** If a legacy plan is in  $T_1$  but not  $T_0$ , its feature information is carried backward to create a record in  $T_0$ . This situation is likely due to the plan not having enough subscribers to place it among the top 100 postpaid (or top 10 prepaid) legacy plans in  $T_0$ , but then appeared in  $T_1$  because subscriber counts of other legacy plans dropped.

Table 3 summarizes the actions for the cases described above:

**Table 3. Processing steps to reduce churn of legacy and in-market plans in the transaction data**

Plan	$T_0$	$T_1$	Action
Legacy plan A	In sample	Missing	Carry <b>forward</b> plan A’s feature information to $T_1$
Legacy plan B	Missing	In sample	Carry <b>backward</b> plan B’s feature information to $T_0$
In-market Plan C	In sample	Missing	Carry <b>forward</b> plan C’s feature information to $T_1$

#### ASSIGNING COMPARABLE PLANS TO SERVICE PACKAGES

Sample analysis of plans has revealed that WSPs often launch promotional plans with the same values as existing plans, but at different prices. A secondary processing step of assigning identical plans into a “**service package**” (SP)

<sup>10</sup> If the missing legacy plan is a prepaid plan, then its subscriber counts are compared to the subscriber counts of the top 10 prepaid legacy plan in  $T_1$ . Note that this imputation procedure for subscriber counts is applied whenever plans are carried forward or backward.

is necessary in the compilation of the index. A service package is comprised of two or more plans which have the same values for the following plan and derived features:

**Table 4. Features used to assign plans into service packages**

<b>Feature name</b>	<b>Description</b>	<b>Provided by WSP?</b>
<b>Geography</b>	the province or territory where a plan is offered	Yes
<b>Retail Brand</b>	the name of the retail brand offering the cellular service plan	Yes
<b>Plan Type</b>	indicator for prepaid or postpaid plans	Yes
<b>Data</b>	data allowance, expressed in Gigabytes (GB) per month	Yes
<b>isMinUnlimited</b>	binary variable that indicates if the service package includes unlimited calling minutes	No--derived from Voice feature
<b>isSMSUnlimited</b>	binary variable that indicates if the service package includes unlimited SMS (text) messages	No--Derived from SMS feature
<b>isDataUnlimited</b>	binary variable that indicates if the service package includes unlimited data <sup>11</sup>	No--Derived from Plan Name feature
<b>isDataShareable</b>	binary variable that indicates if the service package includes data that can be shared among the members in a family plan.	No--Derived from Plan Name feature
<b>includesCANUS</b>	binary variable that indicates if the service package includes Canada-US calling	No--Derived from Plan Name feature
<b>Is5G</b>	binary variable that indicates if the service package uses the 5G network	No--Derived from Plan Name feature

Each service package’s price is calculated as the weighted arithmetic average of all its plans’ MRCs. Each plan’s subscriber count is used to compute the weighted price of the service package.

An example of a service package at the plan level is given in Table 5 below. Service package “A” contains all 5G postpaid plans offered by Retail Brand 1 in Province A, having a data allowance of 50 GB, unlimited nationwide talk and text, and unlimited and shareable data. There are two plans in the service package (Plan X and Plan Y) and the weighted average price is \$89.00.

<sup>11</sup> All plans have a fixed integer value for the “Data” feature. This value is the amount of data that a consumer can use in a month before incurring overage fees, or before experiencing reduced network service quality. A plan offering ‘Unlimited data’ means that for any use beyond the monthly data allowance, the provider will reduce the network quality (e.g., data transmission speed) in lieu of charging overage fees. For example, a 10 GB plan with no unlimited data (a value of 0) will incur overage charges if the consumer uses more than 10 GB in a month.

**Table 5. Assigning plan X and plan Y to Service package A**

Service package A		
Plan Name	Plan X	Plan Y
Geography	Province A	Province A
Retail Brand	Retail Brand 1	Retail Brand 1
Plan Type	Postpaid	Postpaid
Data	50 GB	50 GB
isMinUnlimited	1	1
isSMSUnlimited	1	1
isDataUnlimited	1	1
isDataShareable	1	1
includesCANUS	0	0
is5G	1	1
MRC (price)	\$95.00	\$85.00
Subscribers	40,000	60,000

Service packages are the fundamental building blocks on which pure price change is measured. They also serve to classify the data into two distinct observation types:

- 1) **Continuing:** service packages which are in both  $T_0$  and  $T_1$ .
- 2) **Entering:** service packages which are in  $T_1$  but not in  $T_0$ .

Entering service packages are not carried backward to  $T_0$  since they are composed of new plans. Thus,  $T_0$  prices for entering service packages are imputed via hedonic regression.

**Table 6. Actions for different types of service packages**

Service package type	$T_0$	$T_1$	Action
Service package C (Continuing)	In sample	In sample	No action
Service package D (Entering)	Missing	In sample	Impute $T_0$ price for Service package D

The monthly file is comprised mostly of continuing service packages—a small number of service packages enter and exit the sample every month. Most entering or new service packages do not attract sufficiently large numbers of subscribers to be significant in terms of market share. However, some service packages gain larger numbers of new subscribers during certain months (in particular, during select statutory holidays or back to school events) when the retailers usually offer new plans with data allowances that were not previously available in the market.

#### QUALITY ADJUSTMENT FOR NEW PLANS AND SERVICE PACKAGES

When retailers introduce new plans that offer larger data allowances, additional features such as international calling, or faster transmission speeds, they can quickly become popular with consumers looking to switch from

their current cellular plans. In this context, there is a need to apply quality adjustment methods to account for price change that would otherwise be missed if the index were based on a matched model approach only.<sup>12</sup>

---

## WEB COLLECTED DATA

When a WSP stops advertising a plan that was being priced for a consumer profile, a new plan that fits the profile must be chosen. Out of all the feasible plans, the least expensive is matched with the profile. This action is consistent with a consumer wishing to maximize their consumption of other goods and under a budget constraint. While Canadians might not always subscribe to the cheapest package that meets their needs, it is more reasonable to assume this behavior than to assume that consumers make no effort to minimize their costs. Quality adjustment, following the option cost approach is then used to estimate the price of the new plan in  $T_0$ .<sup>13</sup> In situations where the actual  $T_0$  price of the substituted plan is known (i.e., the substituted plan existed in  $T_0$ ), it is used.

---

## TRANSACTION DATA

To predict  $T_0$  prices of new service packages (in  $T_1$ ), a hedonic imputation approach based on a trained hedonic regression model was taken. Various algorithms were considered for hedonic imputation, including ordinary least squares (OLS), multivariate adaptive regression splines (MARS), and eXtreme Gradient Boosting (XGBoost).

The hedonic regression model is:

$$\ln(MRC_i) = f(X_i) + \varepsilon_i$$

where:

$X_i$  consists of the following explanatory features used to predict the (log-transformed) MRC for service package  $i$ :

Retail Brand, Plan Type, Data, *isMinUnlimited*, *isSMSUnlimited*, *isDataUnlimited*, *isDataShareable*, *includesCANUS*, and *is5G*.

Service packages in the  $T_0$  file, composed of in-market plans only, are used to train a hedonic model in every province or territory to predict  $T_0$  prices of new service packages entering the provincial or territorial market in  $T_1$ . Legacy plans in the  $T_0$  file are not used to train the model because they were not offered in the market in  $T_0$  and thus reflect prices from a period prior to  $T_0$ .<sup>14</sup> Between 30 to 80 services packages were used to train the provincial and territorial models each month.

Once trained, the model is used to predict the  $T_0$  prices of entering service packages in their respective province or territory.

---

<sup>12</sup> Triplett, J. (2004), "Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products", OECD Science, Technology and Industry Working Papers, No. 2004/09, OECD Publishing, Paris, <https://doi.org/10.1787/643587187107>.

<sup>13</sup> This adjustment is based on price per GB (controlled within the retail brand and geography) and it is done when WSPs make significant modifications to a plan's features like offering bonus data.

<sup>14</sup> For example, a postpaid plan that had unlimited talk & text and 4 GB could cost as much as \$75 prior to the [government-mandated 25% reduction in wireless prices](#). Given the dynamic nature of the cellular services market, it makes sense to leave out these old plans from training the model. Currently, one can obtain a 20 GB plan for \$34.



## AGGREGATION

### WEB COLLECTED DATA

Once plan prices for consumer profiles have been collected, and quality adjustments have been applied,  $T_1$  prices are directly compared to  $T_0$  prices to construct price relatives for each consumer profile. The consumer profile price relatives are then aggregated using an unweighted geometric mean to produce a price relative for a WSP in a province or territory.

### TRANSACTION DATA

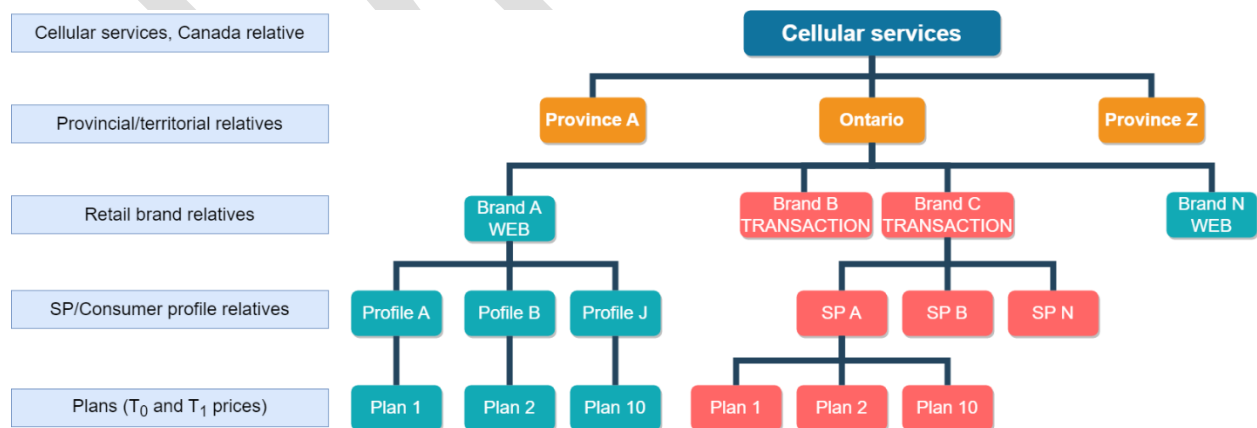
After estimating the  $T_0$  price of all entering service packages, a price relative of every service package is then calculated as a ratio of each service package's  $T_1$  price and  $T_0$  price. Each service package's price relative is weighted by its average revenue share<sup>15</sup> over  $T_0$  and  $T_1$ . Given the availability of revenue information, the Törnqvist index formula was chosen to compute the monthly price change for each retail brand in each province or territory.

### HYBRID INDEXES

The transaction data price relatives are then combined with the web collected data price relatives from the other retail brands in the same province or territory using a weighted geometric mean to produce a “hybrid” provincial or territorial price relative. The resulting provincial or territorial index is a hybrid index because it is compiled using transaction data from WSP's retail brands and manually web collected data from other WSPs' retail brands in the same market. Provincial or territorial retail brand weights are based on the latest revenue data from the AST.

The Canada level price relative is obtained by computing the weighted arithmetic mean of all provincial and territorial relatives. Weights for this aggregation are sourced from the 2022 National Household Final Consumption Expenditure (HFCE) dataset. Given that not all WSPs in the market have agreed to provide the agency with transaction data, the hybrid CSPI will continue to be built on transaction data and web collected data.

Figure 1. Aggregation plan of the hybrid Cellular services price index

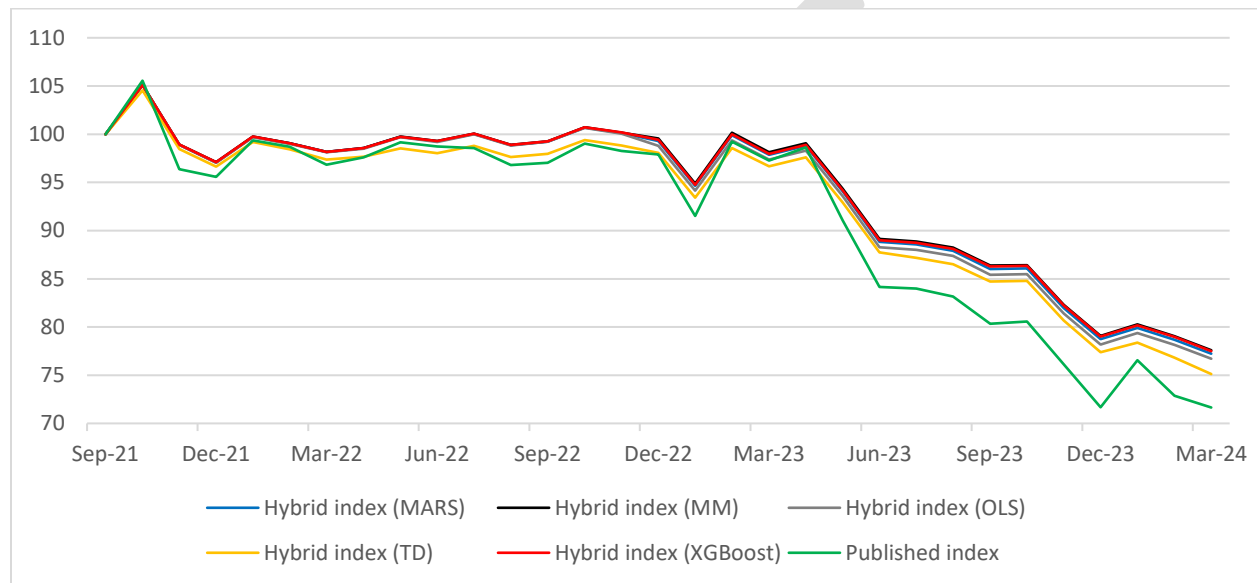


<sup>15</sup> The monthly revenue of a service package is the sum of all monthly revenues for all plans in a service package. A plan's monthly revenue is the product of the plan's subscriber count and its monthly recurring charge.

## RESULTS

As mentioned in the section on quality adjustment, various algorithms were tested when creating the hedonic regression models. The following graph shows various Canada-level Cellular services price indexes differentiated by quality-adjustment method:

**Figure 2. CSPI value by quality adjustment method, Canada**



The published CSPI (green line) is composed solely of web collected data. It is more volatile compared to the other indexes since the composition of the sample is limited to advertised plans on WSP websites. In comparison, the other hybrid indexes display moderate movements due to inclusion of transaction data, which is primarily composed of continuing service packages. As mentioned earlier, new service packages that require quality adjustment usually have fewer subscriber counts so they have a smaller influence on the index's overall movement (outside of peak promotional periods). The three hedonic imputation indexes (OLS, MARS, and XGBoost) performed similarly, but the MARS algorithm is preferred as it produced more accurate predictions, with the lowest Root Squared Mean Error (RMSE) for out-of-sample predictions compared to the other algorithms (see the Appendix for a discussion).<sup>16</sup>

## INTEGRATION TIMELINES AND PLANS FOR OPERATIONALIZATION

The use of transaction data to measure price change for cellular services is unprecedented, so it is vital that the proper methodology is introduced with this hybrid approach. To this end, Statistics Canada is working with price experts, other national statistical organizations, and key stakeholders to ensure that the data and methods used in the calculation of the hybrid CSPI are aligned with international standards and best practices.

<sup>16</sup> For confidentiality reasons, the retail brand level indexes (composed of purely transaction data) are not included. It was found that the Matched Model (MM) index failed to capture significant signals in the market because it did not account for the influx of new service packages during peak promotional periods. As well, the time dummy (TD) index produced unacceptable results, showing large movements that greatly deviated from the rest of the hedonic imputation indexes (OLS, MARS, and XGBoost).

Implementation of this methodology in production is planned over the coming quarters. The following details are being evaluated or are planned and will be elaborated upon in a future document:

- A detailed peer review of the Machine Learning approach in line with the Statistics Canada’s Responsible use of Machine Learning framework to validate the proposed model and approach, training data, impact on production, etc.<sup>17</sup>
- An investigation into the interpretability of Artificial Intelligence methods to help explain how the proposed model predicted a specific price.<sup>18</sup> This activity would also be integrated in the monthly quality assessment process of the data to help analysts understand why the model predicted a specific back price of new service packages.
- A production system that will maintain both the chosen model (hedonic imputation via MARS) in parallel with a simpler model (hedonic imputation via OLS). The ‘out-of-sample’ production performance of both models will be evaluated on a holdout set of  $T_0$  in-market plans not used to train the models.
  - The production process is being developed to handle complex eventualities and may include use of the simpler linear model in production where the complex situation requires a linear model.

## CONCLUSION

Transaction data from WSPs are a comprehensive source of information that can be used to predict prices for a given set of service plan characteristics. Results show that incorporating transaction data into the currently published Cellular services price index had a moderating effect and eased the volatility that was inherent in having a smaller, more limited (web-collected) sample. The addition of transaction data, which is primarily composed of continuing service packages, minimized the effect of quality adjustment on new service packages (and plans) on the CSPI. The hedonic imputation model using MARS was the preferred imputation method to predict prices for new service packages because it had the best performance metrics compared to the other algorithms. Statistics Canada is utilizing additional tools to enhance interpretation of the MARS model. In addition, predictions will be computed using both types of hedonic models (OLS and MARS model) for analytical purposes, and to compute quality metrics on a regular basis for ongoing assessment of model performance. These measures, supported by regular interactions with wireless services providers, will ensure continued production of a high-quality Cellular services price index in the Canadian CPI using a transparent methodology. The incorporation of transaction data into the currently published Cellular services price index will lead to an improvement in the quality of the index.

---

<sup>17</sup> See [Responsible use of machine learning at Statistics Canada \(statcan.gc.ca\)](https://www150.statcan.gc.ca/n1/pub/75-662-x/2019001/article/00001-eng.htm) for more details on the agency framework.

<sup>18</sup> Possible methods include [Individual Conditional Expectation \(ICE\)](#) or Shapley values.

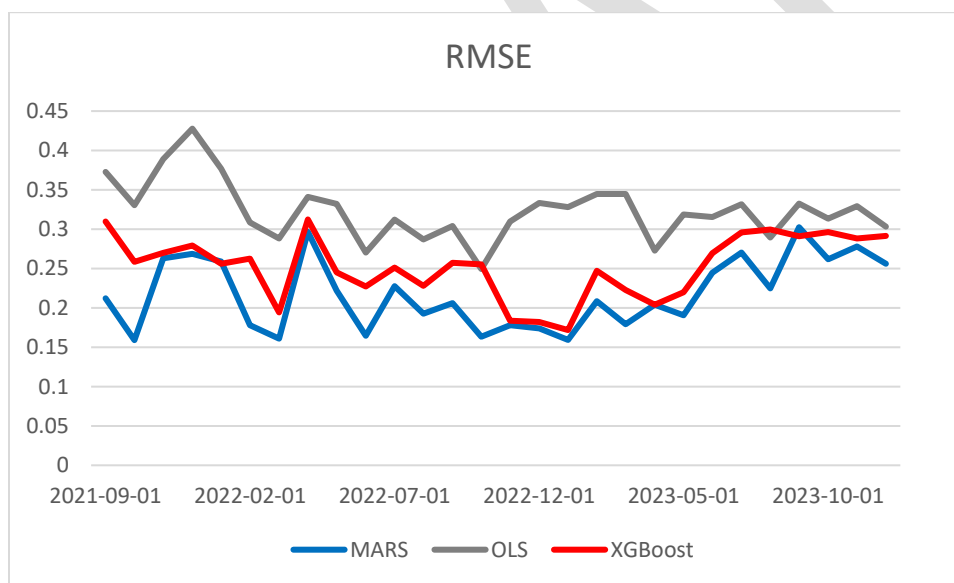
## APPLICATION OF MACHINE LEARNING METHODS FOR HEDONIC QUALITY ADJUSTMENT

Statistics Canada’s proposal to use the MARS algorithm for hedonic quality adjustments was based on analysis of prediction accuracy of various algorithms, including OLS, MARS, and XGBoost. Performance of the models for price prediction was assessed using monthly data from September 2021 to December 2023.

Two repeats of 3-fold cross validation were used to assess model performance. The average prediction error in each month for each algorithm was calculated based on the algorithm’s average prediction error on all six held-out sets.<sup>19</sup>

Figures 3 and 4 below show that the MARS model usually has the lowest RMSE for out-of-sample predictions and the highest  $R^2$  (coefficient of determination) metrics compared to the other algorithms. MARS also has the lowest mean absolute error (MAE) and mean absolute percentage error (MAPE).<sup>20</sup>

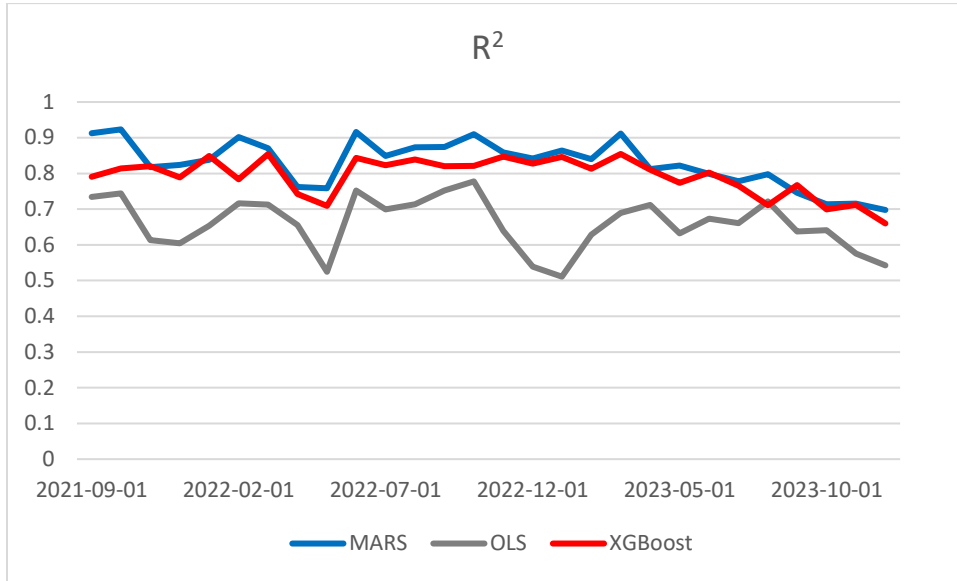
**Figure 3. Monthly RMSE by algorithm**



<sup>19</sup> For each month, the data were randomly divided into 3 equal subsamples or “folds”. A single fold was set aside as the validation data used for the analysis of the out-of-sample prediction error of the model and the remaining two folds were used to train the model. The estimated model was then used to make price predictions for the service packages of the reserved fold and the RMSE, the MAE, the MAPE, and the  $R^2$  metrics were recorded. This subsampling process was then repeated 2 times, with the dataset being split into different folds each time. The metrics from all 6 held-out sets were averaged to estimate the model’s average prediction performance.

<sup>20</sup> Figures 3 and 4 are based on 1 province; however, the rest of the geographies displayed similar results.

Figure 4. Monthly R<sup>2</sup> by algorithm



#### APPLICATION OF THE MARS MODEL

The MARS model is applied separately in each province/territory each month. Transaction data for only one month, comprising on average between 30 and 80 in-market service packages for each province/territory, are used to estimate the hedonic model for that month. Out-of-sample price predictions depend only on that month's hedonic model. As a result, only in-market service packages of plans available in the prediction reference month can affect the prediction.

The regression model predicts the  $T_0$  (log-transformed) MRCs for new service packages using the following explanatory features: Retail Brand, Plan Type, Data allowance, isMinUnlimited, isSMSUnlimited, isDataUnlimited, isDataShareable, includesCANUS, and is5G. Other features (e.g., cellular data transmission speeds (download and upload), geographic coverage, jitter, or latency) are not included as explanatory features primarily because of lack of data. Furthermore, most of these excluded features generally apply to all available plans and they remain unchanged over many months.