INTERNATIONAL MONETARY FUND

IMF Country Report No. 22/303

UNITED ARAB EMIRATES

TECHNICAL ASSISTANCE REPORT—HEDONIC METHODS FOR PRICE INDICES MISSION

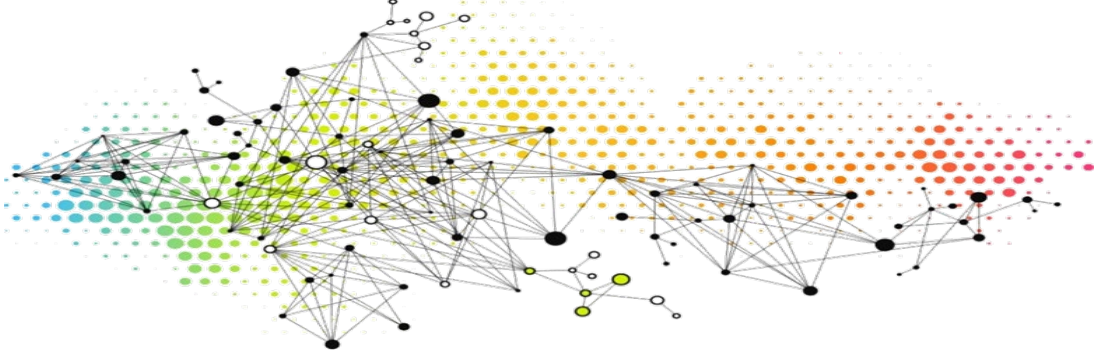
September 2022

This Technical Assistance report on the United Arab Emirates was prepared by a staff team of the International Monetary Fund. It is based on the information available at the time it was completed on January 2022.

Copies of this report are available to the public from

International Monetary Fund • Publication Services
PO Box 92780 • Washington, D.C. 20090
Telephone: (202) 623-7430 • Fax: (202) 623-7201
E-mail: publications@imf.org Web: http://www.imf.org
Price: $18.00 per printed copy

International Monetary Fund
Washington, D.C.



UNITED ARAB EMIRATES

MARCH 2022

REPORT ON HEDONIC METHODS FOR PRICE INDICES MISSION (JANUARY 16–20, 2022)

Prepared By Vanda Guerreiro

CONTENTS

Glossary

COICOP	Classification of Individual Consumption According to Purpose
CPI	Consumer price index
DLD	Dubai Land Department
DSC	Dubai Statistics Center
REPI	Real Estate Price Index
RPPI	Residential Property Price Index
TA	Technical assistance
UAE	United Arab Emirates

SUMMARY OF MISSION OUTCOMES AND PRIORITY RECOMMENDATIONS

1. In response to a request from the Dubai Statistics Center (DSC), a technical assistance (TA) mission took place remotely during January 16–20, 2022 to assist on introducing the use of hedonic methods for quality adjustments in the consumer price index (CPI) and in the Real Estate Price Index (REPI).

2. DSC is willing to improve the current methodology for the compilation of its experimental REPI. Currently the REPI is compiled using stratification with simple averages. Moving from simple averages to hedonic regressions will improve the accuracy of the indicator since it will take into account the quality mix of properties within each stratum.

3. The mission recommended the use of the hedonics time dummy method with a rolling window of 12 months for the compilation of the RPPI. There are a number of hedonic methods which can be used for the RPPI compilation, of those, the mission recommended the use of the hedonics time dummy method with a rolling window of 12 months. This method provides more stable results, i.e., less volatile indices, since it pools one year of data instead of one quarter, and it is particularly recommended when few observations are available. It is widely used for the RPPI compilation and for the CPI compilation with web scraped data. Extended training was provided on this method accompanied by R codes adapted to the Dubai sample data.

4. The REPI is compiled with data obtained from the Dubai Land Department (DLD) available since 2016. These data cover all types of properties namely residential buildings, commercial buildings, and land (residential and commercial). Regardless of obtaining an overall real estate index covering all types of properties (the REPI), a Residential Property Price Index (RPPI) that follows the price trend of the residential buildings should be compiled and published as a stand-alone indicator. The RPPI is a key indicator for the monitoring of financial stability. The same applies to commercial properties, stand-alone indices should be published for each type of activity. Finally, separate indices are needed for residential land and commercial land (by type of activity).

5. The mission created R codes to proceed with the data cleaning, analysis, preparation of data for processing and modeling, based on sample data provided by DSC. The analysis and outlier's detection were performed by strata (sectors) for each month using price per square meter. Different options for the outliers should be tried out by changing the value of the interquartile multiplier.

6. DSC is calculating flow weights using transaction data to aggregate the strata indices into higher level indices. The mission recommended to update the weights annually with transaction data of the previous year, or of the previous three years and keep those stable for the year (Laspeyres-type index).

7. The release of the new RPPI should include the publication of a technical note and/or a

methodology paper. A draft technical note was provided by the mission. The release should be publicized in the media to inform all potential users, namely real estate developers, financial institutions, households, and all institutional users such as tax office, Central Bank, National accounts staff, etc.

8. Web scraping is being increasingly used for the compilation of CPI sub-indices to better account with quality changes in the varieties. The most common products being web scraped are flights, electronics, and clothes. Sub-indices are first calculated with the web scraped data using the time dummy hedonic method with 12-months rolling window. The sub-indices compiled with the web scraped data, for each web site, are then aggregated. The two sub-indices; with web scraped data and in-person data; are also aggregated. Turnover weights of the companies/shops included in each are used for these aggregations. Turnover data can be obtained from the business register and/or tax offices and it is often available for the national accounts.

To support progress in the above work areas, the mission recommended a detailed action plan with the following priority recommendations to make headway in improving the REPI CPI:

Table 1. Priority Recommendations

Target Date	Priority Recommendation	Responsible Institutions
Subject to DSC Management evaluation and approval	Use hedonic methods for the REPI compilation	**DSC**
Every year	Update the weights annually	**DSC**
Continuous	Web scrape all available data for each product, three times per month.	**DSC**

Further details on the priority recommendations and the related actions/milestones can be found in the action plan under *Detailed Technical Assessment and Recommendations*.

DETAILED TECHNICAL ASSESSMENT AND RECOMMENDATIONS

Priority	Action/Milestone	Target Completion Date
H	Compile experimental RPPI with time dummy hedonic method	February 2022
M	Compile other experimental Real Estate Indices with time dummy hedonic method	March 2022
H	Draft a new methodology paper/technical note to inform users and managers of the changes	April 2022
M	Publish new REPI	Subject to DSC Management evaluation and approval
M	Meet with real estate data providers to improve data quality and understand variables in the current dataset	February 2022
H	Begin web scraping data on mobile phones, TV, and other electronic products	April 2022
M	Compile experimental for the web scraped products with time dummy hedonic method	Subject to DSC Management evaluation and approval
M	Investigate data sources on companies' turnover (weights) for the web scraped products	December 2022
	Draft a new methodology paper/technical note to inform users and managers of the changes	Subject to DSC Management evaluation and approval
M	Publish new CPI	Subject to DSC Management evaluation and approval

A. THE REPI

9. DSC is willing to improve the current methodology use for the compilation of its experimental REPI. Currently the REPI is compiled using stratification with simple averages. Moving from simple averages to hedonic regressions will improve the accuracy of the indicator since it will take into account the quality mix of properties within each stratum.

10. Main users and key stakeholders should be engaged in the process. DSC should meet with the Central Bank, with DLD and other main stakeholders to share the current methodology and the future.

11. The REPI is compiled with data obtained from the DLD available since 2016. These data cover all types of properties namely residential buildings, commercial buildings, and land (residential and commercial). The available main variables are the value, if a flat or a house, the location (area and sector), the size in square meters, the existence of a balcony and of a car parking, information of whether the property is existing or off plan, and procedure type. The meaning of some variables and its instances remains unclear, for example for procedure type, thus the mission recommended DSC to meet with the DLD to clarify. After a table as shown in figure 1 should be completed for the data. The

meeting with DLD should also aim at improving future data by completing the built year, that has high number of missing values, and to extent the data to other meaningful variables.

Figure 1. Data Structure

Data structure	Description	Type
Period	Month, quarter, and year	Categorical
Id	Unique identifier of the dwelling	Numerical
Year	Year	Categorical
Month	Month	Categorical
Dwelling_Type	Type of property	Qualitative/Categorical
Price	Registered price of the dwelling	Numerical
Floor Area	Total floor square meters (including common areas)	Numerical
Neighborhood_Affluence	*See text just below this table**	Numerical
Year Built	Interval within the dwelling was built	Categorical
BER	Building Energy Rating	Categorical
County	County where the dwelling is located	Qualitative/Categorical
Region	Region where the dwelling is located	Qualitative/Categorical
Status	New/Existing	Categorical
Central_Heating	Heating system existing in the dwelling	Qualitative/Categorical
Neighborhood_Type	Rural / Urban	Categorical
Building_Levels	Number of story's in the dwelling	Categorical

Source: RPPI Practical Compilation Guide.

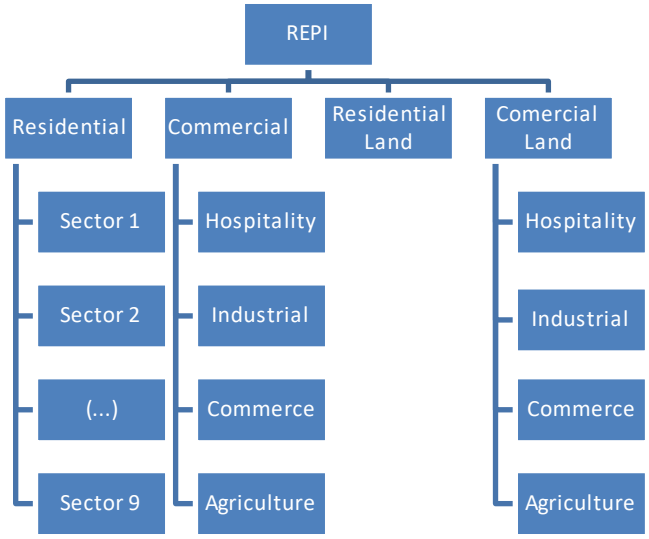
12. An overall real estate price index, the REPI, is compiled covering all property types. The current structure is as follows:

Figure 2. Current Structure of the REPI

Source: Dubai Statistics Center.

13. The mission recommended to change the current structure to have an RPPI. Regardless of obtaining an overall real estate index covering all types of properties, a RPPI that follows the price trend of the residential buildings should be published as a stand-alone indicator. This is a key indicator for the monitoring of financial stability. The same applies to commercial properties, stand-alone indices should be published for each type of activity. Finally, separate indices are needed for residential land and commercial land (by type of activity). The mission suggested the following structure:

Figure 3. Recommended Structure of the REPI



14. The mission created R codes to proceed with the data cleaning, analysis, preparation of data for processing and modeling, based on sample data provided by DSC. From the initial number of observations, the following steps are put in place to clean the data:

- Removing duplicates
- Removing observations with missing values in the variables that are used in the model.
- Identification and removal of outliers
- Creating categories for the number of bedrooms.

The analysis and outlier's detection were performed by strata (sectors) for each month using price per square meter. Different options for the outliers should be tried out by changing the value of the interquartile multiplier. Figure 4 is a print screen of the R code where that experiment should be made, for reference.

Figure 4. Print Screen of the R code, for Reference

```
251
252 # parameters of the outliers - keep stable at least for a year
253
254 BaseCalcul3 <- BaseCalcul2
255 BaseCalcul3$FlgOut <- ifelse ( BaseCalcul3$P_sqm <
256                                (BaseCalcul3$`P_sqm : 1st Qu.`- 0.5 * BaseCalcul3$dif.interqrtl.P_sqm ) |
257
258                                BaseCalcul3$P_sqm >
259                                (BaseCalcul3$`P_sqm : 3rd Qu.` + 0.5 * BaseCalcul3$dif.interqrtl.P_sqm ),
260
261                              1, 0)
262
```

15. DSC is calculating flow weights using transaction data to aggregate the strata indices into higher level indices. The mission recommended to update the weights annually with transaction data of the previous year, or of the previous three years and keep those stable for the year (Laspeyres-type index).

16. DSC aims at improving the RPPI methodology by using hedonic methods instead of simple averages. There are a number of hedonic methods which can be used for the RPPI compilation, of those, the mission recommended the use of the hedonics time dummy method with a rolling window of 12 months. This method provides more stable results, i.e., less volatile indices, since it pools one year of data instead of one quarter, and it is particularly recommended when few observations are available. It is widely used for the RPPI compilation and for the CPI compilation with web scraped data. Extended training was provided on this method accompanied by R codes adapted to the Dubai sample data.

17. The RPPI can be revised up to two quarters prior the reference date. The revisions policy should be clearly publicized on the web site and on the methodology note/paper. The RPPIs are revised in many countries relying on data from administrative sources in order to include transactions pertaining previous quarters registered with delay.

18. The release of the new RPPI should include the publication of a technical note and/or a methodology paper. A draft technical note was provided by the mission. The release should be publicized in the media to inform all potential users, namely real estate developers, financial institutions, households, and all institutional users such as tax office, Central Bank, National accounts staff, etc.

Recommended Actions:

- Meet with main users and key stakeholders to share the current methodology and the future.
- Meet with the DLD to clarify current data and to improve future data collection
- Adapt the current structure of the REPI to have an RPPI, commercial properties indices and residential land index and commercial land indices.
- Experiment with different options for the outliers by changing the value of the interquartile multiplier.
- Update the weights annually and keep those stable for the year.
- Use of the hedonics time dummy method with a rolling window of 12 months
- Release of the new RPPI with a technical note and/or a methodology paper

B. Web Scraping

19. Using web scraped data holds the promise of improving the quality of the CPI. Web scraped data has several advantages over traditional price collection, namely by providing price information over longer periods of time rather than on just one day per month; it is a better source of information for the inclusion of new items in the; it can reduce the administrative burden on retailers and save costs on price collection; and it is expected to increase both the retailer and the item coverage. In addition, improvements in the efficiency of the process of index production are

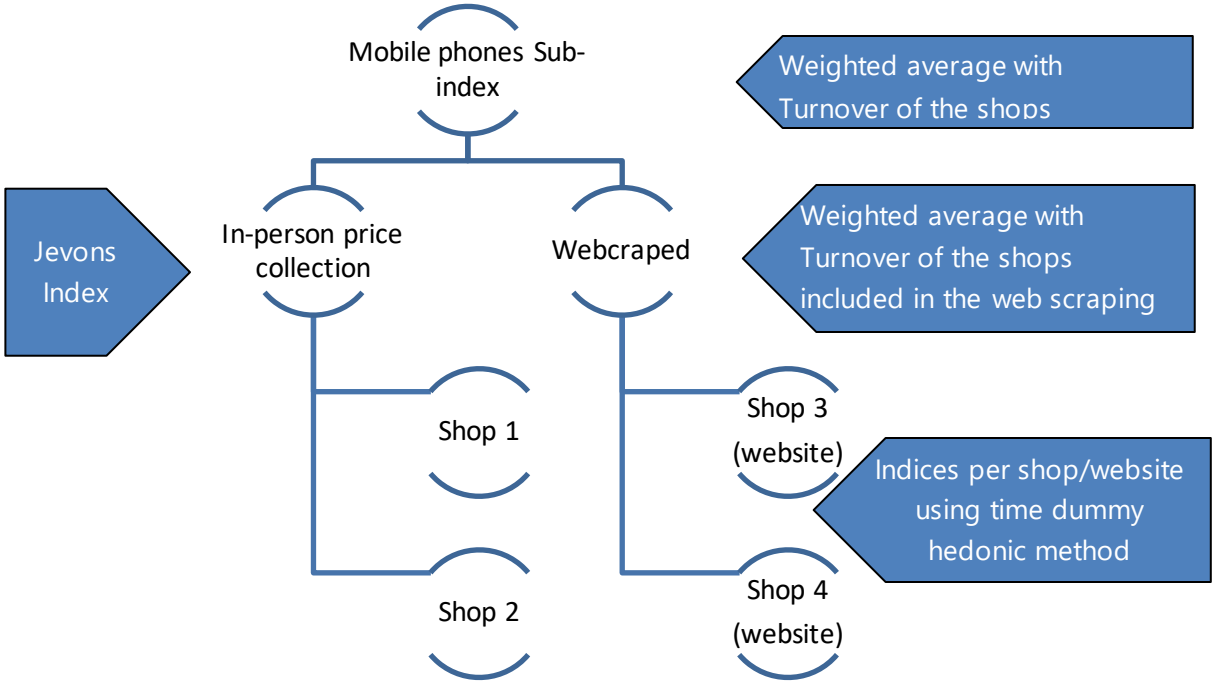
also expected due to the higher level of automatization that can be achieved. Once in place, web scraped data can be less time consuming for a better result.

20. Web scraping is being increasingly used for the compilation of CPI sub-indices to better account with quality changes in the varieties. The most common products being web scraped are flights, electronics, and clothes. The web scraping activity should be performed once per week during the first three weeks of the month. After, all data pertaining one month, and one product is joined in one dataset. All available data should be retrieved, i.e., all varieties and all characteristics of each variety. More than one web site should be scraped for each product.

21. The web scraped data needs to clean for outliers and missing values before the compilation of the indices. The methodology for the cleaning and analysis are performed as before for the real estate indices, for one month of data, for each product. Each characteristic of the varieties corresponds to a variable/column, most of those are categorical variables and will become dummies during calculation. In many cases, when there are a high number of instances (more than five) categories are created as before for number of beds.

22. Sub-indices are first calculated with the web scraped data using the time dummy hedonic method with 12-months rolling window. The recommended method is the same for the reasons mentioned before. In Figure 5, a breakdown of the CPI for mobile phones is shown to exemplify how web scrape is integrated in the CPI. A sub-index is compiled for each web site and each product. The sub-indices compiled with the web scraped, for each web site, are then aggregated (Laspeyres-type) using weights obtained from turnover. Several physical shops will still be subject to in-person price collection for the sampled varieties. a CPI sub-index for the in-in-person price collection data source is compiled as usual (Jevons). The two sub-indices with web scraped data and in-person data are aggregated using turnover weights of the companies/shops included in each. As possible, use the turnover as close as possible to the product concept. For example, for a company that sales a wide range of consumer goods try to have turnover for electronic products only. The turnover concept/coverage must be the same for all companies that are being aggregated. If one only includes technology while others include all turnover, the result will be biased towards the others. This type of turnover data can be obtained from the business register and/or tax offices and it is often available for the national accounts. An excel file was provided exemplifying how this calculation should be made.

Figure 5. Integration of Web Scraping into CPI



23. Data collection should be done in both online and physical shops if the prices or the varieties. In this case you will need to estimate the volume of turnover for online and offline. Otherwise, the in-person price collection can be replaced by web scraping.

24. The introduction of this new form of data collection does not imply a series break and should be communicated to users by updating all metadata and methodology documents.

Recommended Actions:

- Perform web scraping once per week during the first three weeks of the month.
- Retrieve all available data should, i.e., all varieties and all characteristics of each variety.
- Web scrape more than one web site for each product.
- Use the time dummy hedonic method with 12-months rolling window.
- Use turnover weights to aggregate product sub-indices by web site and further to aggregate the web scraped sub-indices with the in-person price collection subindices, by product.
- Investigate potential data sources for turnover with national accounts, tax authorities, business statistics, etc.
- Keep data collection in both online and physical shops if the prices or the varieties are different.

C. Officials Met During the Mission

Name	Institution
Maryam Al-Mulla	DSC
Nusaiba Al Marzooqi	DSC
Adnan Alafari	DSC

Rola Rashad	DSC
AbdelHamid AbdulHadi	DSC
Sherif Bayoumi	DSC
Maryam AlMarri	DSC
Amro Mohammed	DSC
Mohamid Elfateh	DSC