Data-driven analysis of central bank digital currency (CBDC) projects drivers

Toshiko Matsui and Daniel Perez

Department of Computing, Imperial College London, United Kingdom
{t.matsui19, daniel.perez}@imperial.ac.uk

Abstract. In this paper, we use a variety of machine learning methods to quantify the extent to which economic and technological factors are predictive of the progression of Central Bank Digital Currencies (CBDC) within a country, using as our measure of this progression the CBDC project index (CBDCPI). By extracting and aggregating cross country data provided by several international organisations, we find that the financial development index is the most important feature for our model, followed by the GDP per capita and an index of the voice and accountability of the country's population. Our results are consistent with previous qualitative research which finds that countries with a high degree of financial development or digital infrastructure have more developed CBDC projects. Further, we obtain robust results when predicting the CBDCPI at different points in time.

Keywords: central bank digital currency (CBDC) · digital currency · CBDC project index · machine learning · multilayer perceptron · random forest.

1 Introduction

Recent advances in financial technology and distributed ledgers [16, 20] have paved the way to the extensive use of digital currencies. Although the advancement of these currencies came from private initiatives such as Bitcoin [19], Ethereum [26], and Libra [12, 10], researchers and policymakers are contemplating whether central banks can also issue their own digital currencies, usually referred to as *central bank digital currency* (CBDC).

There has been a great deal of discussion about the implications of introducing CBDCs. Although the concept of a CBDC has existed for quite a long time [25], the stance toward whether central banks should introduce them has changed drastically over the past year. Initially the focus of central banks was on systemic implications [6] but several factors deriving from the benefits of the digital money have recently motivated central banks to issue CBDCs. This trend has further been fueled by the declining use of cash due to the growth of cashless payments, the possible introduction of global stablecoins [15] and the Covid-19 pandemic [2].

In fact, a great number of central banks are undertaking extensive work on CBDC [8], several of which have issued research or statements on the related motivations, architectures, risks, and benefits. For instance, Boar et al. [8] refer to the observed shift to intensive practical development from conceptual research found in emerging markets, driven by stronger motivations than those of advanced economy central banks. In practice, several central banks issued their CBDCs between August and December 2020 (see Section 2.2). Further, a few central banks are aiming to issue their CBDCs in the next few years, which is attracting a lot of attention [3]; for example, the beginning of this year saw the closing of the ECB digital euro consultation with record level of public feedback [13]. This move is consistent with the observation that the introduction of CBDCs can present significant innovations in money and banking history [14].

However, despite the great amount of analysis conducted regarding the important questions surrounding CBDCs, relatively little quantitative analysis has been undertaken especially on the drivers of the CBDC projects. The previous research include the potential risk and benefits of introducing CBDC, and quantitatively, the welfare gains of bringing CBDC into the economy [11]. Of the relevant research that exists in this vein, [3] suggests, by taking an ordered probit approach [18] with a comprehensive cross country database, that the majority of the CBDC projects are found in digitised economies with a high capacity for innovation. They conclude that some of the potential drivers of CBDC development are related to factors affecting a country's digital infrastructure, innovation capacity, institutional quality, development and financial inclusion, public interest in CBDCs, and cross-border transactions.

This study examines the economic and institutional drivers of CBDC projects by applying machine learning techniques to the related variables obtained from official sources that are available for a wide cross section of countries. Our primary objective is to improve the understanding of the dominant drivers for CBDCs and the factors that increase the possibly of a country to accelerate this effort. We use the CBDC project index (CBDCPI) [3] as our objective variable and factors affecting a country's digital or technological capability and government effectiveness as independent variables, in order to reduce the problem to identifying the independent variables with the most predictive power. To accomplish this, we utilise machine learning techniques to predict the CBDCPI and pick the most important variables for our model.

We compare two types of classifiers that are able to learn non-linear functions: a multilayer perceptron (MLP) [21, 22] and a random forest [9]. In the experiment, we find random forest performs better than MLP, and that the financial development index [23] is the most important feature for our model, followed by the GDP per capita and the voice and accountability, when explaining the CBDCPI drivers for August 2020. This concurs with [3], which concluded that more developed CBDC projects can be found in countries with higher financial development index, digital infrastructure, GDP, and institutional characteristics. As a robustness check, we have performed the same analysis with full and aggregated data and with December 2020 CBDCPI. Results are broadly consistent, although there were some minor changes in the ranks of important features.

This paper is structured as follows. Section 2 describes the preliminaries including the overview of CBDCs and the CBDC project index (CBDCPI). We subsequently present the empirical models and data in Section 3, then results in Section 4. We conclude in Section 5.

2 Preliminaries

In this section, we provide an overview of the key concepts from central bank digital currency (CBDC) relevant to this paper.

2.1 Central bank digital currency (CBDC)

A central bank digital currency (CBDC) is the digital form of the fiat currency of a particular nation (or region). It differs from virtual currency and cryptocurrency, as CBDC is issued by the state and possesses the legal tender status declared by the government [17]. Examples include the Digital Currency/Electronic Payments (DC/EP) by China's central bank and e-krona by the central bank of Sweden.

The introduction of CBDCs is receiving more attention than ever before. Although the concept of a CBDC was already proposed decades ago [25], recent IT progress and its application to the financial industry have motivated central banks and academics to study the risks and benefits of making CBDC accessible to the general public [5, 7], as presented in Table 1 [1].

Table 1. Benefits and challenges of CBDCs

Advantages	**Disadvantages**
Low cost of cash	Higher run risk
Enhances financial inclusion	Commercial bank disintermediation
Stabilises the payment system	Enhances currency substitution
Faster monetary policy transmission	Risk and cost for central banks

Further, attitudes towards whether CBDCs should be issued by central banks have changed drastically over the past year. This derived from the diminishing use of cash due to the rise of cashless payments, the possible entry of global stablecoins [15] and the Covid-19 pandemic. Specifically, social distancing measures and public concerns that cash may transmit the Covid-19 virus and novel government-to-person payment schemes have further fueled the shift toward digital payments, and may act as a driver of CBDC projects [2].

In fact, CBDCs have gained global attention, not only within central bank communities but also by the public. Fig.1 charts the Google search interest over time[1]. This shows that the current interest in CBDC is increasing, reaching a level almost as high as Bitcoin's during its price spike of 2017.

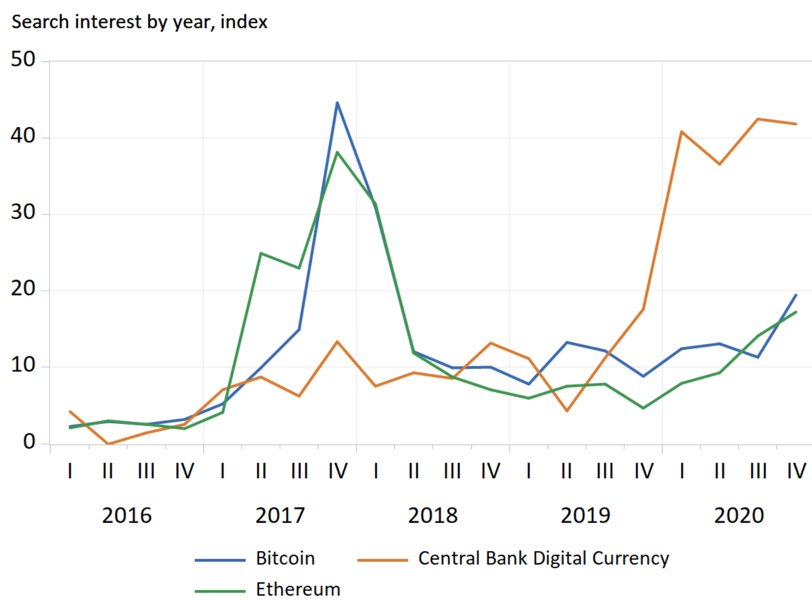

Fig. 1. Google Trends over time

However, a majority of CBDCs are still in research or pilot stage, although a survey in early 2020 showed more than 80% of central banks were studying the subject [24]. Further, the drivers of the CBDC projects are yet to be thoroughly investigated [3].

2.2 CBDC project index (CBDCPI)

The *CBDC project index* (CBDCPI) was firstly proposed by Auer et al. [3] to measure the central bank's progress toward the development of a retail or

[1] We took 12-week moving average of Google Trends search results.

wholesale CBDC. CBDCPI represents publicly announced work by central banks on CBDC related projects. The index takes a value between 0 and 4 defined as follows:

0 - No announced project
1 - Public research studies
2 - Ongoing or completed pilot
3 - Live CBDC

There are two sub-indices, one for retail and one for wholesale CBDC projects. Wholesale CBDC is devised as a new instrument for settlement between financial organisations, whereas retail CBDC aims to replace cash with the properties of central bank liability. The overall index for a country is the maximum of these two sub-indices.

According to the dataset provided by the online annex of [3], there was no country with index 3 (live CBDC) as of August 2020 but by the end of 2020, seven countries and jurisdictions, including Bahamas, Canada, Switzerland, France, have shifted to 3.[2] This clearly reflects the recent trend and discussion that have brought the analysis about digital money and CBDC to the fore [4].

3 Data and Methodology

Our main goal is to understand better the main drivers for CBDCs and the factors that makes a country more or less likely to push this effort. Using the CBDC project index (CBDCPI) as our objective variable and factors affecting a country's digital or technological capability and government effectiveness as explanatory variables, we can reduce this to finding the explanatory variables with the most predictive power. To achieve this, we leverage machine learning techniques to predict the CBDCPI and extract the most important variables for our model. We use the rest of this section to describe our data and provide details about our machine-learning based methodology.

3.1 Data

We extract our dataset from the World Bank, the International Monetary Fund (IMF), and the data source of a BIS working paper [3]. After having processed and cleaned the data, it ended up containing 16 variables, including data from 2000 to 2019, with more than 170 countries and jurisdictions. The Table 2 lists the observed variables included in our analysis.

[2] The updated CBDC projects status is available in an online annex of [3] (See https://www.bis.org/publ/work880.htm). The information is said to have been collected through desk research and with the help of contacts at several individual central banks.

Table 2. Table of observed variables

Variable	Description	Source
Digital infrastructure		
Mobile subscriptions	Subscriptions to a mobile telephone service (per 100 ppl.)	WB
Secure Internet	The number of secure Internet servers (per 1 million ppl.)	WB
Broadband subscription	Fixed broadband subscriptions (per 100 ppl.)	WB
indiv. Internet use	Individuals using the Internet (% of population)	WB
Development and financial inclusion		
Account ownership	Account ownership at a financial institution or with a mobile-money-service provider (% of population aged 15+)	WB
FD index	Indices that summarise the degree of developments of financial institutions and financial markets	IMF
GDP per capita	GDP divided by midyear population (USD)	WB
Institutional characteristics		
Government effectiveness	Quality of public services, policy implementation etc.	WB
Regulatory quality	Ability to formulate and implement sound policies etc.	WB
Voice and accountability	Extent of citizens' participation and freedom expression	WB
Innovation environment		
Access to electricity	Access to electricity (% of population)	WB
Demographic characteristics		
% of people over 65	Total population 65 years of age or older	WB
Cross-border transactions		
Trade (% of GDP)	Sum of exports and imports (% of GDP)	WB

Source: World Bank (DataBank), IMF (working paper [23]).
Dependent variable, CBDCPI, was obtained from BIS (online annex of [3]).

Our variables can be divided into the following categories: i) digital infrastructure, ii) development and financial inclusion, iii) institutional characteristics, iv) innovation environment, v) demographic characteristics, and vi) cross-border transactions. Each variable contains several years of data ranging from 2000 to 2019. Although our dataset on financial development index holds data about several financial development sub-indices in terms of their depth, access, and efficiency, we only include the top-level index[3]. We perform our analysis both using the *full data* and an aggregated version of our data. For the *aggregated data*, we average each variable over the period 2014-2019, subject to data availability.

The CBDCPI has 176 observations, one per country, each taking the value of 0, 1, 2, or 3, as described in Section 2.2. We obtain the CDCPI for December 2020, in addition to its August 2020 value from previous research [3]. The major difference between these two variables is that, as of August 2020, there were

[3] See Financial Development Index Database by IMF for more information. (https://data.imf.org/?sk=f8032e80-b36c-43b1-ac26-493c5b1cd33b)

Table 3. Countries with a CBDCPI of 3 as of December 2020

Country	**Overall** (Dec 20)	**Overall** (Aug 20)	**Retail** (Dec 20)	**Wholesale** (Dec 20)
Bahamas	3	2	3	0
Canada	3	2	1	2
Switzerland	3	1	1	2
Euro area (ECB)	3	2	1	2
France	3	2	1	2
Japan	3	2	1	2
South Africa	3	1	1	2

Source: Online annex of [3].

no countries with a live CBDC (i.e. no index with value 3) while there were 7 of them with a live CBDC by December 2020. We show the countries with a CDCPI of 3 as of December 2020 in Table 3.

3.2 Methodology

Instead of the ordered probit approach [18] applied in [3], we model the problem as a classification task where the goal is to predict the CBDCPI given the set of input variables described above. Given that the CBDCPI is a value between 0 and 3, it is easy to model as a categorical variable.

We settle on a random forest as our primary model, as it is known to be able to learn complex non-linear functions while being interpretable enough to extract the most important input variables [9]. To obtain a point of comparison for the predictions of our random forest, we also train a multilayer perceptron [21, 22] on the same task. However, given that multilayer perceptrons are not interpretable enough to understand the most important features, we only use these results for comparison.

We utilise this methodology with the full version and the aggregated version of the data to predict the CBDCPI both in August and December 2020.

4 Results

This section presents the results we obtained by training the models described above on our dataset.

Before starting our training process, we preprocess the data to filter out lacking data. We first remove all the countries for which the CBDCPI is not available, as well as the countries for which one or more of the observed variables is not available (e.g. do not have a single year of data). When a country is missing a year for a particular variable, we use the previous year to fill for it (e.g. if the GDP per capita is available for 2018 but not 2019, we set it to the value of 2019). After this filtering process, we obtain a final list of 145 countries with only 6 countries having a CBDCPI of 3 as of December 2020; unfortunately, the Euro Area was not included in the final dataset as we did not have the financial development index data for it. Further, we obtain a total of 13 variables for our aggregated data and 135 variables for the full data.

Then, we randomly split our dataset in two equal splits for training and testing. We then tune the hyper-parameters of our two models. We note that given the small size of our dataset, we use the full data instead of having a separate cross-validation set. We find that our random forest works best with a total of 100 estimators. For our multilayer perceptron, we use two layers, the first one with a number of neurons equal to the number of features and the second one a fixed size of 10 neurons. Finally, we train the two models on our training data and evaluate them on our test data. We present the accuracy of the two models in Table 4. Overall, our model performs better with the full data rather than its aggregated version.

Table 4. Accuracy of different classifiers on full and aggregated data

		Full data		**Aggregated data**	
	Classifier	Train	Test	Train	Test
Aug	**MLP**	1.0	0.79	0.99	0.74
	Random Forest	1.0	0.78	1.0	0.77
Dec	**MLP**	1.0	0.67	1.0	0.62
	Random Forest	1.0	0.78	1.0	0.68

Next, we use the features extracted by our random forest to understand better what the potential drivers of CBDC are. We use the aggregated August 2020 data to compare with [3] and see if the random forest achieves similar results as to the drivers for CBDCs. We find that the financial development index [23] is by far the most important feature for our model, followed by the GDP per capita and the voice and accountability, when explaining the CBDCPI drivers for August 2020. This is consistent with [3], stating that the CBDC projects to be more developed where there is higher financial development index, GDP, digital infrastructure, and institutional characteristics such as govenment effectiveness and voice and accountability. We summarize the top 5 features and their importance for random forest in Table 5.

For robustness check, we conduct the same analysis with December 2020 CBDCPI data as an objective variable. The results are almost consistent with the August CBDCPI data, aside from the fact that the aging rate additionally accounts for the December 2020 data – random forest performing much better and allows to extract the most important features used for classification. The most

Table 5. Most important features for the random forest (Aug 2020, aggregated)

Feature	**Importance**
Financial development index	0.165
GDP per capita	0.098
Voice and accountability	0.095
Broadband subscriptions (per 100 people)	0.093
Government effectiveness	0.087

Table 6. Most important features for the random forest (Dec 2020, aggregated)

Feature	**Importance**
% of people over 65	0.134
Financial development index	0.109
Mobile cellular subscriptions (per 100 people)	0.101
GDP per capita	0.090
Voice and accountability	0.089

significant features are: mean 65+, financial development index, mobile cellular subscription, GDP per capita, and voice and accountability, showing that the main features predicted as important are proved to be important with December 2020 CBDCPI data, as suggested in Table 6.

5 Conclusion

In this paper, by applying a variety of machine learning methods used to learn complex non-linear functions for comprehensive cross country data, we investigated the importance of each economic and technological factors in predicting the progression of Central Bank Digital Currencies (CBDC) project within a country, using as our measure of this advancement the CBDC project index (CBDCPI). We found that a financial development index is the most important feature for our model, followed by the GDP per capita and an index of the voice and accountability of the country's population. Additionally, we confirmed that our results are in accordance with previous qualitative research which finds that countries with a high degree of financial development or digital infrastructure have more advanced CBDC projects. Moreover, we achieved robust results when examining the CBDCPI at different points in time.

References

1. Adrian, T., Mancini-Griffoli, T.: Central bank digital currencies: 4 questions and answers, IMF Blog, 12 December 2019

2. Auer, R., Cornelli, G., Frost, J.: Covid-19, cash and the future of payments. BIS Bulletin 3 (Apr 2020)
3. Auer, R., Cornelli, G., Frost, J.: Rise of the central bank digital currencies: drivers, approaches and technologies. BIS working paper, No. 880 (Aug 2020)
4. Bank for International Settlement: Central bank group to assess potential cases for central bank digital currencies. BIS press release (21 Jan 2020) https://www.bis.org/press/p200121.htm
5. Bank for International Settlement: Central bank digital currencies. CPMI, Markets committee papers, No. 174 (Mar 2018)
6. Barontini, C., Holden, H.: Proceeding with caution - a survey on central bank digital currency. BIS papers, No. 101 (Jan 2019)
7. Bindseil, U.: Tiered CBDC and the financial system. Working paper series 2351, European Central Bank (Jan 2020), https://ideas.repec.org/p/ecb/ecbwps/20202351.html
8. Boar, C., Holden, H., Wadsworth, A.: Impending arrival - a sequel to the survey on central bank digital currency. BIS papers, No. 107 (Jan 2020)
9. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001), https://link.springer.com/article/10.1023%2FA%3A1010933404324
10. Brühl, V.: Libra - a differentiated view on facebook's virtual currency project. CFS working paper series 633, Frankfurt a. M. (2019), http://hdl.handle.net/10419/206412
11. Davoodalhosseini, S.M.R.: Central bank digital currency and monetary policy. Staff working papers, Bank of Canada (2018), https://EconPapers.repec.org/RePEc:bca:bocawp:18-36
12. Diem Association: Diem white paper (2020), https://www.diem.com/en-us/white-paper
13. European Central Bank: ECB digital euro consultation ends with record level of public feedback. ECB press release (13 Jan 2021) https://www.ecb.europa.eu/press/pr/date/2021/html/ecb.pr210113 ec9929f446.en.html
14. Fernández-Villaverde, J., Sanches, D., Schilling, L., Uhlig, H.: Central bank digital currency: Central banking for all? Review of Economic Dynamics (2020), http://www.sciencedirect.com/science/article/pii/S1094202520301150
15. Financial Stability Board: Regulation, supervision and oversight of "global stablecoin" arrangements. Report, Financial Stability Board (Oct 2020)
16. von zur Gathen, J.: CryptoSchool. Springer-Verlag Berlin Heidelberg (2015), https://link.springer.com/book/10.1007%2F978-3-662-48425-8
17. Grym, A., Heikkinen, P., Kauko, K., Takala, K.: Central bank digital currency. Tech. rep., Banque de France (2017)
18. McKelvey, R.D., Zavoina, W.: A statistical model for the analysis of ordinal level dependent variables. The Journal of Mathematical Sociology **4**(1), 103–120 (1975), https://www.tandfonline.com/doi/abs/10.1080/0022250X.1975.9989847
19. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2009), http://www.bitcoin.org/bitcoin.pdf
20. Narayanan, A., Bonneau, J., Felten, E., Miller, A., Goldfeder, S.: Bitcoin and cryptocurrency technologies: A comprehensive introduction. Princeton University Press, USA (2016)
21. Rosenblatt, F.: Principles of neurodynamics: Perceptions and the theory of brain mechanism. Spartan Books (1961)
22. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation, p. 318–362. MIT Press, Cambridge, MA, USA (1986)

23. Svirydzenka, K.: Introducing a new broad-based index of financial development. IMF working papers, No. 15 (2016)
24. The Economist: Will central-bank digital currencies break the banking system? Tech. rep., ISSN 0013-0613 (Dec 2020)
25. Tobin, J.: The case for preserving regulatory distinctions. Proceedings of the Economic Policy Symposium p. 167–83 (1987)
26. Wood, G., et al.: Ethereum: A secure decentralised generalised transaction ledger. Ethereum project yellow paper **151**(2014), 1–32 (2014)

A Appendix

This annex gives additional tables, regression results and figures to complement the paper. See main text for further discussion.

A.1 CBDC projects status

Below shows the part of the updated project score of global CBDC development efforts, relating to [3] (as of December 2020)[4]. Note that only the countries with index of 3 (live CBDC) and 2 (pilot) as of December 2020 are listed here.

Country	Overall*	Overall (Aug 20)	Retail*	Wholesale*
Bahamas	3	2	3	0
Canada	3	2	1	2
Switzerland	3	1	1	2
Euro area (ECB)	3	2	1	2
France	3	2	1	2
Japan	3	2	1	2
South Africa	3	1	1	2
United Arab Emirates	2	2	0	2
Australia	2	1	1	1
China	2	2	2	0
Ecuador	2	2	2	0
Eastern Caribbean	2	2	2	0
United Kingdom	2	2	1	1
Hong Kong	2	2	0	2
Indonesia	2	1	1	1
India	2	0	1	1
South Korea	2	2	2	0
Saudi Arabia	2	2	0	2
Sweden	2	2	2	0
Singapore	2	2	0	2
Swaziland	2	1	1	1
Thailand	2	2	0	2
Ukraine	2	2	2	0
Uruguay	2	2	2	0

*As of December 2020.

[4] The dataset includes all projects announced as of 1 December 2020. For more information, see https://www.bis.org/publ/work880.htm.

A.2 Top 10 features for the random forest classifier with aggregated data

Table 7 and 8 give the 10 most important independent variables for the random forest classifier with aggregated data (data averaged over the period 2014–19, subject to data availability), with August 2020 and December 2020 CBDCPI data as an objective variable, respectively.

Table 7. Most important features for the random forest classifier (Aug 2020)

Feature	Importance
Financial Development Index	0.165
GDP per capita	0.098
Voice and accountability	0.095
Broadband subscriptions (per 100 people)	0.093
Government effectiveness	0.087
% of people over 65	0.082
Individuals using the Internet (% of population)	0.080
Trade (% of GDP)	0.078
Secure Internet servers (per 1 million people)	0.072
Regulatory quality	0.068

Table 8. Most important features for the random forest classifier (Dec 2020)

Feature	Importance
% of people over 65	0.134
Financial Development Index	0.109
Mobile cellular subscriptions (per 100 people)	0.101
GDP per capita	0.090
Voice and accountability	0.089
Secure Internet servers (per 1 million people)	0.086
Individuals using the Internet (% of population)	0.080
Government effectiveness	0.080
Broadband subscriptions (per 100 people)	0.076
Trade (% of GDP)	0.067

A.3 Top 10 features for the random forest classifier with full data

Table 9 and 10 show the 10 most important index for the random forest classifier with full data, with August 2020 and December 2020 CBDCPI data as an objective variable, respectively.

Table 9. Most important features for the random forest classifier (Aug 2020)

Feature	**Importance**
Financial Development Index	0.052
Government effectiveness [YR2019]	0.033
Government effectiveness [YR2018]	0.020
Broadband subscriptions (per 100 people) [YR2017]	0.020
Individuals using the Internet (% of population) [YR2012]	0.020
Individuals using the Internet (% of population) [YR2015]	0.020
GDP per capita [YR2016]	0.018
Government effectiveness [YR2015]	0.016
Mobile cellular subscriptions (per 100 people) [YR2016]	0.015
% of people over 65 [YR1990]	0.014

Table 10. Most important features for the random forest classifier (Dec 2020)

Feature	**Importance**
Financial Development Index	0.037
% of people over 65 [YR1990]	0.035
Mobile cellular subscriptions (per 100 people) [YR2019]	0.025
% of people over 65 [YR2018]	0.023
% of people over 65 [YR2017]	0.021
% of people over 65 [YR2000]	0.020
Government effectiveness [YR2017]	0.020
% of people over 65 [YR2015]	0.020
% of people over 65 [YR2013]	0.019
Mobile cellular subscriptions (per 100 people) [YR2015]	0.018