



MINISTRY OF
FINANCE



INFORMATION TECHNOLOGY CENTER
OF CUSTOM, TAXATION AND FINANCE



MONGOLIA

POVERTY AND INEQUALITY DURING COVID-19 USING BIG DATA.

MAY 2021

HHH 65
ДAA 330
M-692

The United Nations Development Programme is the leading United Nations organisation fighting to end the injustice of poverty, inequality, and climate change. Working with our broad network of experts and partners in 170 countries, we help nations to build integrated, lasting solutions for people and planet.

General disclaimers:

This is a publication of United Nations Development Programme. The designation of the materials in this publication does not imply the expression of any opinion whatsoever on the part of UNDP, NSO or the government of Mongolia.

The content and views expressed in this publication are those of the authors and do not necessarily reflect the views or policies or carry the endorsement of the co-publishing organisations. The publishing organisation does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. By making any designation of or reference to a particular territory or geographic area, or by using the term "country" in this document, UNDP, NSO and the government do not intend to make any judgments as to the legal or other status of any territory or area.

© **2021 United Nations Development Programme.**

All rights reserved. Published in 2021.

UNDP encourages printing or copying information exclusively for personal and non-commercial use with proper acknowledgment of UNDP. Users are restricted from reselling, redistributing or creating derivative works for commercial purposes without the express, written consent of UNDP.

Authors: Otgontugs Banzragch¹, Manlaibaatar Zagdbazar², Uyanga Gankhuyag³, Delgernaran Tumurtogoo⁴, Amarbal Avirmed⁵, Davaajargal Davaatseren⁶, Undral Lkhagva⁷, Ouyntsetseg Mashir⁸, Batbileg Tumor⁹, Ganbayar Javkhlan¹⁰, and Aruinbayar Galbat¹¹.

Also available online at:

<http://mn.undp.org>

For enquiries, please contact:

United Nations Development Programme

UN House, United Nations Street 14

Sukhbaatar District, 14201,

Ulaanbaatar, Telephone: +976-11-327585

ISBN 978-9919-25-163-5

¹ Professor, Economics Department, National University of Mongolia otgontugs@num.edu.mn

² Researcher, Economic Research Institute of Mongolia manlaibaatar@eri.mn

³ Economist, UNDP Bangkok Regional Hub, uyanga.gankhuyag@undp.org

⁴ Former Economist, UNDP Mongolia, delgernaran.tumurtogoo@undp.org

⁵ Director, Population and Social Statistics Department, National Statistics Office of Mongolia, amarbal@nso.mn

⁶ Statistician at National Statistics Office of Mongolia, davaajargal@nso.mn

⁷ Statistician at National Statistics Office of Mongolia, undral@nso.mn

⁸ Senior Statistician at National Statistics Office of Mongolia, oyuntsetseg@nso.mn

⁹ Director, Tax Data Center, Ministry of Finance, batbileg@itc.gov.mn

¹⁰ Senior economist, Ministry of Finance of Mongolia, ganbayar_j@mof.gov.mn

¹¹ Senior economist, Tax Data Center, Ministry of Finance, ariunbat@gmail.com

FOREWORD

The COVID-19 pandemic has escalated into health and socio-economic crises globally and in Mongolia. The impact of the pandemic has depressed household incomes, increased poverty, and exacerbate inequalities. Thus the response actions must be focused on not only responding to the pandemic more effectively, but also be geared towards longer term recovery.

To provide timely, reliable, and the most relevant data to decision-makers in this pandemic, UNDP and NSO collaborated with the National University of Mongolia (NUM), the Ministry of Finance, and Information Technology Center for Custom, Taxation, and Finance to conduct this pilot research to examine the impact of COVID-19 on consumption, poverty, and inequality in Mongolia in 2020 by using big data – the data generated by the Value-Added Tax (VAT) e-system. This study provided an opportunity to understand changes in household spending in Mongolia, thus allowing to make alternative estimates of poverty and inequality. Most importantly, it showed how VAT data can be used to track changes in spending, poverty, and inequality much more frequently than was possible until now.

Having and analyzing such fast-moving data is critical for policy makers to rapidly see the impact of decisions and make adjustments – which is especially important during this fast-evolving crisis. This pilot study can also be useful for other countries seeking ways to analyze poverty, inequality, and the impacts of shocks when opportunities for traditional data collection are severely restricted.

We hope that the study and suggested methodology will accelerate further applications of big data in Mongolia and overall contribute to a more effective recovery from the impacts of the COVID-19 pandemic on the lives of Mongolians.



Elaine Conkievich

UNDP Mongolia Resident
Representative



Batdavaa Batmunkh

Chairperson of National Statistics Office
of Mongolia

EXECUTIVE SUMMARY

This analysis shows how household expenditures, as well as poverty and inequality of household spending, have changed in Mongolia during 2020, the first year of the COVID-19 period by using big data. This is the first time the big data generated from value-added tax (VAT) records is being used for research purposes in the country.

We find that on average, households' spending increased during the pandemic year of 2020, while poverty and inequality of spending slightly declined (the poverty headcount rate declined by 4.8 percentage points and the Gini coefficient by 0.026 points in 2020 compared with 2019).

However, the big data, which allows monthly tracking of these indicators, shows that household spending had large swings during the year, with decreases in the first and the fourth quarter of 2020, when the strictest lockdowns were instituted in the capital city Ulaanbaatar, at times extended to other cities and provinces. Spending of the poorest two quintiles and the richest quintile was particularly affected by the pandemic. Urban households' spending has declined the most during lockdowns, while rural households' spending was affected more in the first half of the year, perhaps due to the inability to export the livestock produce.

The shock of the pandemic on household spending, especially that of the poorest 40 percent of households (as measured by expenditure), was softened by an economic stimulus package introduced in April 2020. The stimulus spending included a large social protection component, most of which is spent on increased universal child allowance. However, the sustainability of fiscal outlays of this magnitude in the future is unclear.

The big data-based estimates of consumption, poverty, and inequality differ from the official estimates based on household surveys, due to the differences in methodology. However, used in combination with an innovative approach used in this study, big data provides a new avenue for tracking changes in spending, as well as poverty and inequality based on expenditures, on a much more frequent basis than was possible until now, allowing to examine how these indicators respond to policy changes. The design of this research also shows that data constructed from restricted public sector big data can be used as a new source for empirical and policy research without compromising the privacy and confidentiality of data.

ACKNOWLEDGEMENTS

We would like to express our deepest appreciation to all those who contributed to developing and finalizing this publication. Special gratitude goes to Amarbal Avirmed, Director of the Population and Social Statistics Department of the National Statistics Office of Mongolia, and Nashida Sattar, UNDP Mongolia Deputy Resident Representative, under whose leadership this study championing the use of big data was conducted. We thank Uyanga Gankhuyag, Economist at UNDP Bangkok Regional Hub, for conceptualizing this research idea, providing substantive guidance, and contributing to the analysis throughout the project, to Delgernaran Tumurtogoo, Economist at UNDP Mongolia, for managing the project, to Dr. Balazs Horvath, UNDP Asia Pacific Regional Senior Economic and Strategic Advisor and Dr. Haniza Khalid, UNDP Malaysia Economist for peer review and valuable comments. We also thank Buyandelger Ulziikhuu, M&E Analyst at UNDP Mongolia for providing the support in the project implementation. This research would not have been possible without the crucial support and cooperation of the holders of big data – the Value-Added Tax data. In this regard, we express sincere appreciation to Batbileg Tumor, Director of the Information and Technology Center of Customs, Taxation and Finance, and Ganbayar Javkhlan, Senior Economist at the Ministry of Finance.

We are grateful to Dr. Otgontugs Banzragch, Professor, Economics Department, National University of Mongolia, Dr. Manlaibaatar Zagdbaatar, Researcher, Economic Research Institute of Mongolia, for conducting this innovative analysis and writing the study report. We greatly appreciate the methodological guidance provided, as well as data processing and analytical work carried out by government researchers: Oyuntsetseg Mashir, Senior Statistician, Davaajargal Davaatseren, Statistician, and Undral Lkhagva, Statistician at the National Statistical Office; Aruinbayar Galbat, Data Scientist, Information and Technology Center of Customs, Taxation and Finance, thanks to whom this research project was possible - whilst respecting the confidentiality and privacy of data.

TABLE OF CONTENTS

Foreword	iii
Executive Summary	iv
Acknowledgements.....	v
SECTION 1. Introduction	1
SECTION 2. Background on big data	4
SECTION 3. Descriptive statistics	7
SECTION 4. Findings.....	9
4.1. Findings on expenditure dynamics during COVID-19.....	9
4.1.1. Expenditure by male-headed and female-headed households.....	12
4.1.2. Expenditure by rural-urban areas	13
4.1.3. Expenditure by the number of children in the household.....	14
4.1.4. Expenditure by quintiles.....	16
4.1.5. Spending versus income: the savings and loans factor.....	17
4.2. Findings on poverty	18
4.3. Findings on inequality.....	21
4.3.1. Measures of inequality using indices.....	21
4.3.2. Measures of inequality using distribution.....	24
SECTION 5. Conclusions.....	27
5.1. Conclusions on research findings	27
5.2. Recommendations on big data use	29
Annexes	33
6.1. Methodology for processing big data	33
6.2. Second-stage data processing	37
6.3. Reweighting and adjustment for bias.....	37
6.4. Estimation methodologies.....	38
6.4.1. Estimation of consumption (expenditure).....	39
6.4.2. Estimation of poverty.....	43
6.4.3. Estimation of inequality	45

SECTION 1. INTRODUCTION

We do not know the full impact of the pandemic for Mongolia yet, but we know it is colossal. According to the National Statistics Office of Mongolia (NSO), GDP fell by 5.3 percent in 2020 (GDP per capita fell by 6.3 percent) (NSO, 2021). This is a massive decline from the trend growth rate, considering that the average growth rate of GDP in the preceding decade 2010-2019 was 7.8 percent (and GDP per capita growth rate averaged 5.6 percent).

Compared with most other countries globally and in the Asia Pacific region, the spread of the pandemic in Mongolia has been much more contained - as of 27 January 2021, 1,667 people were infected with COVID-19, while 1,256 recovered and 4 deceased. This was due to strict lockdown measures adopted by the Government of Mongolia early in the pandemic, the widespread use of masks, generally high level of compliance of the population with government measures and border control measures¹².

Until March 2021, there were three periods of strict lockdowns, covering Ulaanbaatar city, and extending at times to secondary cities and provincial centers. The first lockdown began on 27th January 2020 and had been relaxed gradually by 1st May 2020. The Government of Mongolia introduced strict measures of social distancing, banned public gatherings, closed public spaces, limited public transportation, closed all educational institutions including kindergartens, schools, colleges, and universities, and required wearing masks. International travel was banned¹³.

While for most of 2020, all cases of infection were “captured” at the border through a quarantine system, due to lax adherence to protocols, the local transmission of COVID-19 started in earnest on 11 November 2020, which prompted the government to impose the second lockdown for 60 days through 11th January 2021. The State Emergency Commission (SEC) banned all businesses from operating except those in 13 vital sectors such as hospitals, power plants, grocery shops, petrol stations, and pharmacies. The authorities limited the city transportation, banned public gatherings and spaces, and closed educational institutions, major trading places, and markets. Only employees of the 13 priority sectors were allowed to travel to and from their place of work. Domestic and international travel continued to be severely restricted. To mitigate the spread of the virus, the government organized random and targeted surveillance testing at various sites.

¹² However, since the local transmission of COVID-19 started in November 2020, the spread of the pandemic has accelerated rapidly, reaching 17,823 confirmed cases and 29 deaths as of 14 April 2021.

¹³ International travel by air has been limited to a few flights per month by the national flag carrier airline, largely bringing Mongolians stranded abroad back into the country. Land borders were also closed, with limited numbers of passengers and truckers allowed at a time.

The third lockdown started on 12th February 2021 on the eve of Lunar New Year and continued till 23rd February 2021. Consequently, the lockdowns have been relaxed and the spread of infections accelerated. However, our study uses only the data for 2020, so the period of the third lockdown is outside the scope of our analysis.

To deal with the economic fallout of COVID-19, the government of Mongolia took a series of measures.¹⁴ On 27th March 2020, the State Emergency Commission (SEC) started to implement a comprehensive set of fiscal measures to protect vulnerable households and businesses and to support the economy with 5.1 trillion MNT (equivalent to 13.7 percent of GDP in 2019, or 1.79 billion USD in current prices and the current exchange rate¹⁵) and measures such as raising monthly child allowances, unemployment benefits, and food stamps; tax exemptions on several imported food and medical items and types of equipment; and tax exemptions on income tax as well as suspension of corporate income tax and social security contributions until the end of September 2020. These measures, while providing a much-needed injection into the economy, will have further pushed Mongolia's debt into an unsustainable territory due to the country's increasing burden of public external debt.

Social protection measures were an important part of the fiscal stimulus. To put the size of social benefits into perspective, the annualized spending to meet the announced commitments on social benefits would be equivalent to 1.7 trillion MNT (about 34 percent of the total announced fiscal stimulus), whereas multiplied by 9 months from April to December 2020, it would run up to an estimated 1.3 trillion MNT (about 26 percent of the announced fiscal stimulus). The largest of these social benefits was the increase in universal child allowance¹⁶, which accounted for nearly 80 percent of the social protection part of the COVID-19 fiscal stimulus. In April 2020, the monthly child cash transfer was increased from 20,000 MNT to 30,000 MNT and then in May, it increased the benefit further to 100,000 MNT per child, per month (US\$35). The government also increased monthly food stamps from 18,000 MNT per poor household to 36,000 MNT. Moreover, before the COVID-19 spread, in November 2019 the Government decided to increase public employee's wages in 2020. Figure 1 illustrates total government spending on social protection in 2020.¹⁷

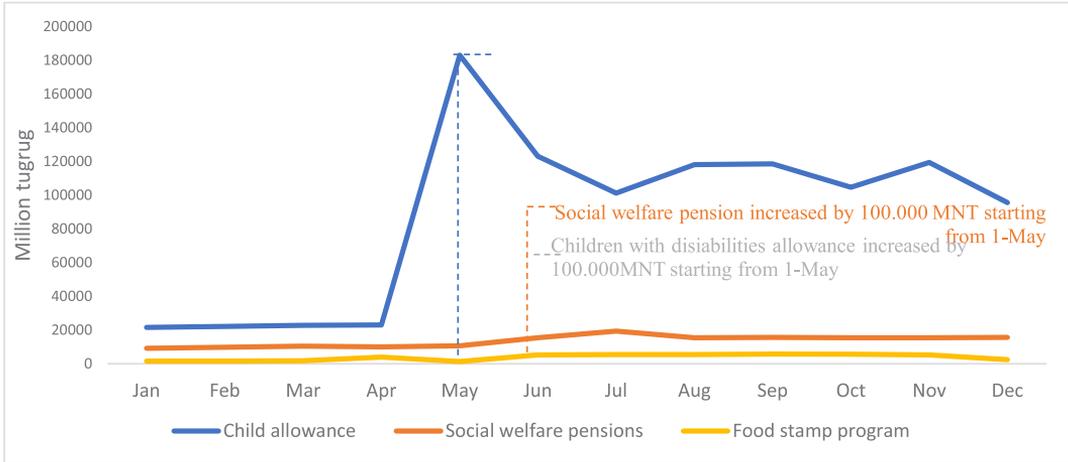
¹⁴ On 11th March 2020, the Bank of Mongolia reduced the policy rate from 11 to 10 percent, and reduced the MNT reserve requirement of banks to 8.5 percent, and narrowed the policy rate corridor to ± 1 percent. The lower reserve requirement released MNT 324 billion (0.8 percent of GDP) of additional liquidity in the banking system. On March 18, the BOM and the Financial Regulatory Commission implemented temporary financial forbearance measures on prudential requirements, loan classifications, and restructuring standards.

¹⁵ 1USD=2850MNT

¹⁶ Prior to COVID-19 fiscal measures, Mongolia has had a universal child allowance, but of a much smaller amount (20,000 Tg per month, equivalent to US\$7 per month).

¹⁷ Child allowance is given for children under 18 years old. Following the increase in child allowance from 30,000 Tg to 100,000 Tg in May 2020 (about US\$35 per month), households with children retroactively received the increment of child allowance for April (70,000 = 100,000 – 30,000). Thus, the budget spending on child allowance in May reflects both 100,000 Tg per child for May plus 70,000 Tg per child retroactive for April. After May 2020, the spending on child benefits has stabilized.

Figure 1. Government spending on social assistance during COVID-19 in 2020



Source: Based on NSO (2020)

Consequently, despite a large decline in real GDP in the first 9 months of 2020, the final real consumption increased by 71 percent year-on-year (NSO, 2020), in large part due to the government stimulus package. However, for the year as a whole, the increase in consumption was only 2.7 percent due to a large decline in the last quarter of 2020 (NSO, 2021).

In this study, we aim to examine the impact of COVID-19 on consumption, poverty, and inequality in Mongolia in 2020 using big data generated from Value Added Tax (VAT) records. This provides an opportunity to analyze consumption, poverty, and inequality – and the impact of the pandemic on them - rapidly and with much greater frequency than traditional household surveys allow. Such granular analysis also allows policy makers to better understand the impact of decisions on social protection and public health on poverty and inequality. Further, such a methodology can potentially be replicated in other countries seeking ways to analyze poverty, inequality, and the impact of shocks at a time when opportunities for traditional data collection are severely restricted.

Consequently, researchers from the NSO, UNDP Offices in Mongolia and the Bangkok Regional Hub, the National University of Mongolia (NUM), and data scientists from the Custom, Taxation, Finance Information System Technology Center (referred to in this report as the Tax Data Center) formed a research team to conduct a study using this data.

Our study seeks to answer the following questions:

1. How did the total consumption and categories of consumption change during the COVID-19 period in Mongolia?

2. What are the differences in consumption depending on the household head's gender, location, and income levels over the COVID-19 period in Mongolia?
3. How did poverty and inequality in terms of expenditure changed over the COVID-19 period at the household and individual level?
4. Is there any difference in estimations of inequality using data from household surveys and big data?
5. How to develop a weighting method between the VAT and HSES data samples to make the matched data representative for the population for further estimates?

In this study, consumption is approximated by expenditure generated from the VAT data.

The rest of this paper is organized as follows. Section 2 sets the background concerning the emerging big data in the country. Section 3 provides descriptive statistics. Section 4 presents findings on expenditure, poverty, and inequality during the COVID-19 period. Section 5 concludes. Further information on data processing and methodologies for estimation of expenditure, poverty, and inequality is provided in the Annex.

SECTION 2. BACKGROUND ON BIG DATA

Big data is commonly defined as data that is big in terms of volume (a large quantity of data), variety (multiple types of data and unstructured data), and velocity (the speed at which data is created). In recent years, the tendency of using non-publicly available data, including big data, in economics research increased significantly (Einav and Levin, 2014).

The VAT records in Mongolia meet this definition of big data – they are automatically generated through the process of consumers and businesses making transactions. Every minute, consumers scan their E-receipts into the system. In 2020 alone, 931 million E-receipts were printed in total.

The VAT in Mongolia is assessed at 10 percent and is the largest source of tax revenues, accounting for about 25 percent of tax revenues. The Parliament of Mongolia adopted a new VAT tax law in July 2015. While increasing the threshold to be registered as a VAT payer¹⁸, the law required all businesses with incomes above this threshold to use the Point of Sale (POS) machines to register all their sales in a unified system. The law also introduced a lottery system and a 20 percent VAT refund¹⁹ to final consumers – payers of VAT, creating an incentive for consumers to demand POS machine-generated receipts from retailers.

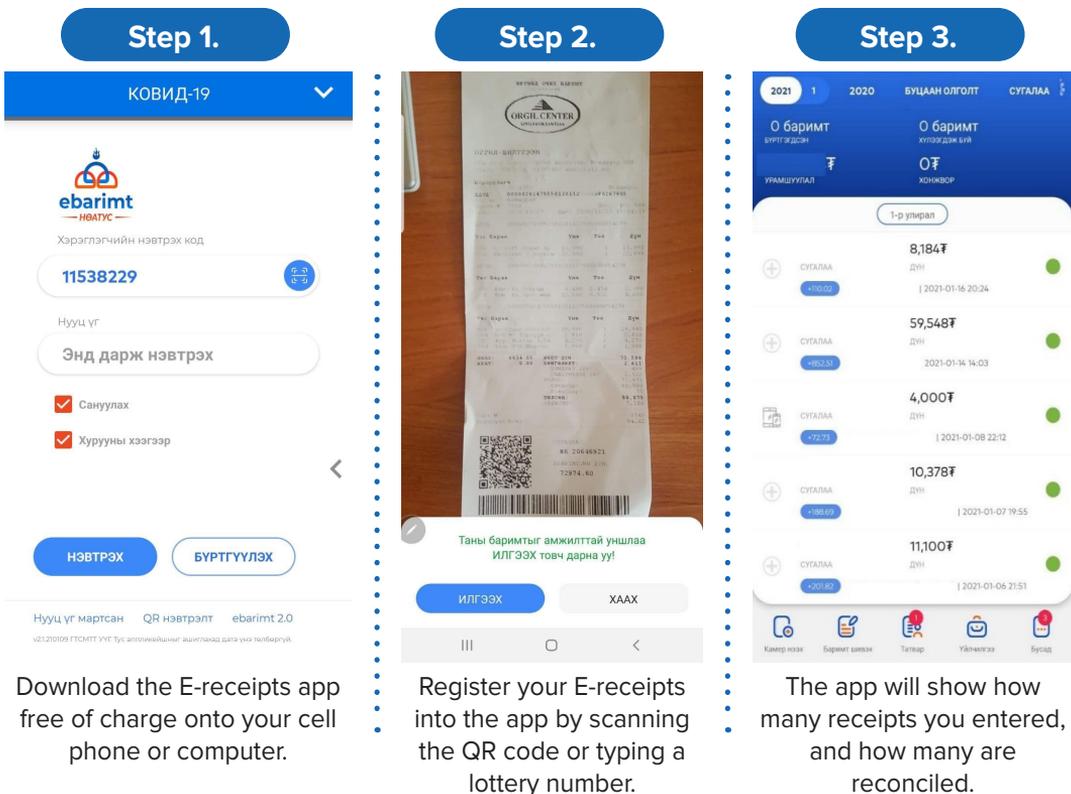
¹⁸ According to the VAT Law of Mongolia of 2015, legal entities are responsible to pay VAT when their total annual sales exceed 50 million MNT (or 18,868 USD at the current exchange rate, December 2020 rate). Before 2015, the threshold level was 10 million MNT for the total sales, while the VAT rate was the same at 10 percent.

¹⁹ 20 percent of the 10-percent VAT, which is equal to 2 percent of purchases.

In tandem with the law, the government created a unified system called the E-receipts system, to which all POS machines in the country are connected. Once a POS registers a sale, the E-receipts system sends a unique QR code to the POS machine. Consumers are encouraged to download an app through which they can scan receipts with the QR codes to register their purchase. Once consumers scan their QR codes into the app, the system reconciles consumers' records with that of the sellers, generating two records for every transaction.²⁰

When the E-receipts system was introduced, 20 percent of the paid VAT had been refunded to consumers - initially at the end of each year, and since 2020, at the end of every quarter. Consumers receive the refund automatically into their registered bank accounts. In addition, the government also organizes an online lottery every quarter with multiple cash prizes ranging from 20,000 MNT to 50 million MNT. The lottery system functions in the following way: the system chooses randomly the winners, notifies about the winning through the E-receipts app, and after several days the prize money is transferred into a winner's bank account.

Figure 2. E-receipt system's registration steps



²⁰ However, the number of the receipts printed (by sellers) has consistently exceeded the number of the receipts scanned (by consumers), although the gap has been declining. For example, in 2017, the number of printed receipts was more than double than that of scanned receipts; in 2017, it was about 60 percent higher and in 2018, about 49 percent higher.

Due to the combination of these legal and administrative measures, the VAT E-receipts system administered by the Tax Data Center rapidly expanded in recent years. Before the implementation of the new law, only about 12,000 POS machines were used in the country. By the end of 2020, the number of POS machines more than quadrupled to 54,677.

As a result, the tax base expanded. Between 2015 and 2019, VAT revenues increased by 136 percent (in nominal terms), with the share of VAT in total tax revenues increasing from 20.5 percent to 25.5 percent (NSO, 2021).

This study allowed us to estimate the rate of use of the E-receipts system using a sample of the 2018 HSE Survey conducted by NSO (referred to as HSES-2018 sample). Table 1 shows that about 43.7 percent of households in the matched HSES-2018 sample were registering their E-receipts into the system in 2016, at the onset of the E-receipts system's establishment; by 2020, this percentage grew to 71.3 percent. Given that the HSES-2018²¹ sample is representative at the national and location levels, these figures should closely approximate what percentage of all households use the E-receipts system.

Moreover, the data shows the uptake of the E-receipts system by location. Based on the matched HSES-2018 sample, in 2020, 80.2 percent of households Ulaanbaatar (the capital city) have used the E-receipts app, registering their expenditure data into the system. Similarly, 76.4 percent of households residing in aimag centres, 65.5 percent of soum centres and, importantly, 47.7 percent of countryside households (mostly nomadic herder families) were registering E-receipts into the system.²²

Table 1. VAT E-receipts system coverage

	Total	Location			
		UB	Aimag center	Soum center	Rural
2016	43.7%	60.1%	48.5%	29.2%	10.9%
2017	51.3%	66.7%	57.4%	38.4%	17.5%
2018	59.5%	73.0%	65.8%	49.5%	27.0%
2019	64.7%	76.3%	70.1%	56.3%	36.4%
2020	71.3%	80.2%	76.4%	65.5%	47.7%

Notes: Shows percent of matched households in the HSES-2018 sample – those whose records were successfully matched with the VAT E-receipts records (see further details in Section 2.2).

Source: Authors' estimate based on analysis of HSES-2018 and VAT E-receipts data

²¹ This and the subsequent paragraph refer to the “matched HSES-2018 sample”. The matched HSES-2018 sample refers a subset of the HSES-2018 sample (13,733 households out of the total 16,646 household surveyed through HSES-2018) whose unique Registration Numbers could be found. For further explanation, see Section 2.2.

²² Aimag centres are provincial centres, whereas soum centres are county/ district centres, equivalent to villages. People living in the countryside means people not living in settlements, i.e., nomadic herders. Ulaanbaatar city and aimag centres are classified as urban locations, whereas soum centres and countryside as rural.

The VAT data is confidential - not available for public use. At the same time, it provides rich information that can be used for decision-making. Beyond the fiscal stimulus measures to address the impact of COVID-19, the VAT data can also be useful for understanding the effectiveness of tax-related measures and helping to guide measures to further expand the tax base.

This study represents the very first time this big data is used in research. As a research team, we have devised a methodology of processing this big data which puts safeguards to ensure adherence to the law and protection of data privacy, while at the same time enabling the researchers to obtain sufficient data to conduct robust analysis and generate insights. This methodology is described in the next section.

SECTION 3. DESCRIPTIVE STATISTICS

In Table 2, we present descriptive statistics comparing households in the HSES-2018 sample with households in the big data sample, which is a subset of the HSES-2018. The data on household characteristics for both sets come from the HSES-2018 data.²³ This comparison shows whether the subset of households in the big data sample differs by important characteristics from households in the HSES-2018 sample.

The full HSES sample and the big data sample are very similar in terms of age of the household head, but different in other characteristics such as gender, education, and employment status of the household head. This means that people registering their transactions in the E-receipts system are more likely to belong to male-headed households, as well as to households headed by people that are employed, better-educated, and slightly younger than heads of households in the general population represented by the HSES-2018 sample.

Table 2. Descriptive statistics of households in the HSES-2018 and the big data sample

	HSES 2018	VAT 2018-2020
Obs.	16,454	4,463
Gender of HH head		
male	75.5%	81.5%
female	24.5%	18.5%
Age of HH head	46.6	44.5
Education level of HH head		
no education	3.6%	1.0%
primary	8.2%	2.8%
basic	14.3%	10.5%

²³ The big data sample by itself does not contain any information on households to which individuals belong, or information on the characteristics of the households.

secondary	24.7%	26.8%
technical and vocational	17.2%	19.7%
technicum	7.3%	7.8%
non-bachelor diploma	7.4%	9.1%
bachelor	14.9%	19.3%
master and doctor	2.3%	3.1%
Employment status of HH head		
employed	67.0%	76.5%
unemployed	6.0%	5.2%
studying	0.4%	0.2%
elder	19.7%	12.2%
disabled	3.3%	2.2%
housework and taking care others	2.1%	2.0%
not working with other reason	1.4%	1.6%

Note: all estimates are weighted.

Source: Authors' estimates

Similarly, we can see from descriptive statistics where incomes and expenditures of households in the big data sample differ from those in the full HSES-2018 sample.²⁴ Table 3 shows that on average, income and expenditure of households in the big data sample are somewhat higher than those in the HSES-2018 data sample. However, the standard deviations are also very large, so there is no statistically significant difference between the two means.

Table 3. Income and expenditure of households in the HSES-2018 sample versus the big data sample

	HSES-2018 sample (n=16454)*	VAT big data sample (n=4463)**
Income per person, MNT per month		
Mean	330,901	344,930
St.dev	(285,792)	(292,027)
Expenditure per person, MNT per month		
Mean	347,617	373,781
St.dev	(262,569)	(265,152)

Source: Authors' estimates

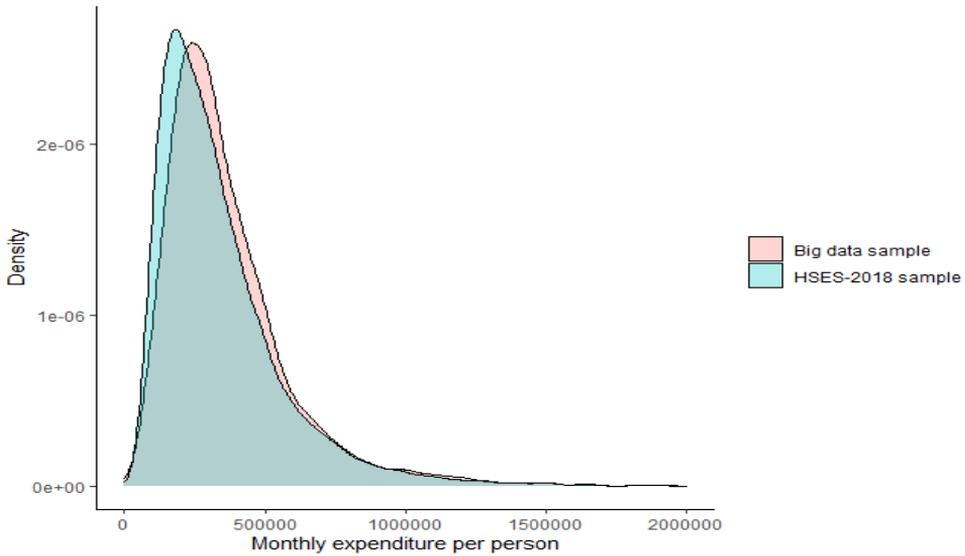
Notes: The means and standard deviations are weighted.

** Weighted by HSES-2018 weights ** Weighted by the big data sample weights*

²⁴ The income and expenditure data shown in this section on descriptive statistics is not taken from the VAT big data. Instead, it shows the incomes and expenditure taken from HSES-2018, for both the full HSES-2018 data set (16,454 households), and its subset – those households which we include in the final big data sample (4,463 households).

The density distribution of expenditure per person of the two data sets shown in Figure 3 also indicates that households in the big data sample have higher expenditure – and therefore, are somewhat wealthier compared to the full HSES-2018 sample.

Figure 3. Distribution of expenditure in the overall HSES-2018 sample versus the subset of the big data sample



Source: Authors' estimates

SECTION 4. FINDINGS²⁵

This section presents findings on how household consumption, poverty, and inequality have changed during the COVID-19 period, using expenditure data generated from the big data sample.

4.1. Findings on expenditure dynamics during COVID-19

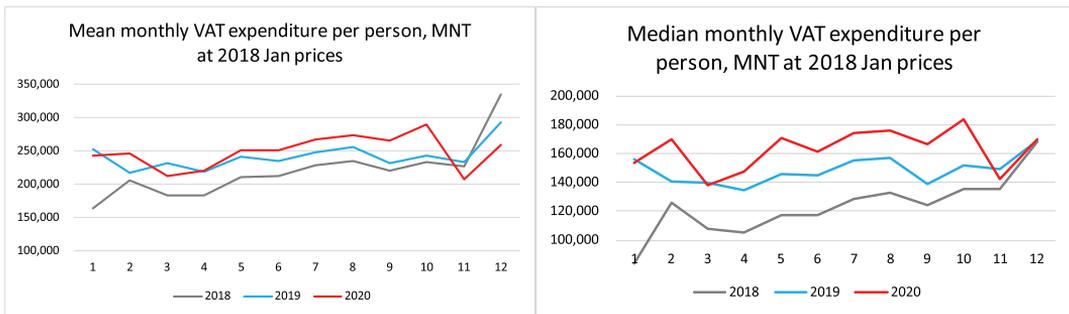
The first of our five research questions was how the total consumption and categories of consumption have changed over the COVID-19 period in Mongolia. We use expenditure data as a proxy for consumption. Average annual expenditure increased by 3 percent in 2020 at constant prices, which is a positive change, but is still significantly lower than the 10 percent increase in 2019. This significant deceleration of expenditure growth is mainly due to the COVID-19 pandemic. Using

²⁵ All expenditure (spending) figures used in this chapter use household-level per-person monthly expenditures adjusted for inflation (base period = January 2018), unless specified otherwise. Also, due to pronounced seasonality in spending in Mongolia, most changes reported are year-on-year changes.

monthly disaggregation, we can see more detailed dynamics of this expenditure trend.

We found that in January 2020, just after COVID-19 struck in the country, mean expenditure fell by 3 percent compared to the January 2019 level. In February 2020, because of the Lunar New Year celebration, even though the government instituted a strict lockdown during this month, mean expenditure grew by 13 percent compared to February 2019 level.²⁶ However, in March 2020, it fell by 9 percent compared with March 2019, revealing the effect of the COVID-19 lockdown. (See Figure 4, section a).

Figure 4. Trends in monthly expenditure in 2018-2020



(a) Mean expenditure

(b) Median expenditure

Due to the announcement of fiscal stimulus measures on 27th March, household expenditures did not decrease in April 2020, compared with April 2019.²⁷ Subsequently, as the child allowance was more than tripled, households started to spend more in May, June, and July 2020 (year-on-year), with expenditure increasing by as much as 20 percent in October 2020 (year-on-year). However, since the second, stricter lockdown started on 11th November 2020, expenditure fell by 11 percent in both November and December (year-on-year).

In terms of median measures of expenditure, changes were similar, but of a smaller magnitude (See Figure 4, section b). In January 2020, median expenditure fell by 2 percent (year-on-year). After an increase in February 2020, due to the Lunar New Year effect, median expenditures declined again, by 1 percent compared with March 2019. The decline in per-person spending observed after the initial COVID shock was reversed starting in April. The year-on-year increase in spending was further amplified in May. For the following months, the expenditure was higher than in the previous year. It was only during the second lockdown which started in November

²⁶ The Lunar New Year in 2019 was in the beginning of February (5-7 February), whereas in 2020, it was at the end of February (24-26 February). Usually, the preparatory spending for the Lunar New Year takes place in about two to four weeks before the holiday. Therefore, there was an increase in Lunar New Year-related spending in January 2019, whereas that in 2020 it occurred in February.

²⁷ Even though households did not receive child allowances until May, the decision on raising child allowances was announced in April, which may have led to households spending more – in advance.

that median monthly expenditure fell by 5 percent year-on-year. In December 2020, the change was zero percent year-on-year. Thus, in November, median expenditures declined by less than mean expenditures, whereas in December, median expenditures rose slightly while mean expenditures continued to decline.

Overall, our findings show that despite the overall decline in economic activity in Mongolia, the year-on-year decline in consumption happened only in some months of 2020 - the same months when strict lockdowns were implemented.²⁸ Except in February, both the first and the second lockdowns have had a large and immediate impact on the decline in consumption.

How did the stimulus measures and payments made to households in mid-April 2020 affect household expenditures? In addition to increasing social benefits, the government also suspended personal income tax payments and social security contributions from April (See Figure 1). Even though households did not receive the increased social benefits until May, households' mean monthly per person expenditure increased compared to the previous month and also year-on-year already starting in April (See Figure 4). This could mean that measures other than social benefits – such as suspension of tax and social security contributions – could already have affected the increase in consumption. It could also mean that households responded to expectations – starting to increase spending even when the increase in social benefits was announced, but not yet distributed. Thus, by April 2020, spending was restored.

The fact that median expenditures declined by less or rose by more in 2020 compared with mean expenditures means that expenditures at the lower end of the distribution (expenditures of poorer households) were less affected by the economic downturn of 2020. For instance, the median consumption did not decline by as much during the second lockdown in November-December. This could mean that the increase in social benefits – which occupy a bigger share of expenditure of poorer and typical households, compared with that of richer households – has been effective in preventing the decline in expenditures of these households. Even though this pattern did exist during the first lockdown of January-March, it was less pronounced; moreover, it should be noted that even before the increase of social benefits in April, Mongolia has had universal child benefits in place, albeit of a smaller magnitude.

Our findings are consistent with empirical evidence from ADB research, where the study concluded that “The short-term policy response of the Government has been significant and pro-poor and should have avoided many of the negative economic effects of the pandemic that we have seen in other countries” (ADB, 2020).

The second research question we aimed to investigate was: “What are the differences in consumption by the household head's gender, the number of children, location, and income levels over the COVID-19 period in Mongolia?”

²⁸ With the exception of February, the month of the Lunar New Year.

Incomes and consumption in Mongolia have a highly pronounced seasonal character. Therefore, to ensure that seasonal variation is taken into account, we have examined and reported year-on-year changes, rather than month-on-month.²⁹ For 2019 and 2020, these consumption figures are also adjusted for inflation to enable comparability with 2018.

4.1.1. Expenditure by male-headed and female-headed households

Interestingly, female-headed households in the big data sample have higher expenditure compared with male-headed households for the entire 2020 period (both mean and median monthly per person expenditure). However, the direction of change in expenditure is largely the same for both male- and female-headed households. Figure 5 shows that in the big data sample, expenditure of female-headed households is both higher, on average, and more dispersed compared with the expenditure of male-headed households. Moreover, the confidence intervals of the two measures do not overlap during most of the 2019-2020 period, unlike in the HSES-2018 data set, indicating that the differences in expenditure are statistically significant for most of the period.

The reasons for this can be the following:

Women in Mongolia are much more likely to have higher education compared with men (61.5 percent compared with 38.5 percent for men) (NSO, 2019). Also, while gender segregation of employment exists by sectors, and women tend to work in lower-paid sectors, they are also widely employed in higher-paid sectors such as financial services and trade. For these reasons, the earning gap between women and men is smaller than that in most other countries at a comparable level of development.

It could also mean that among female-headed households, the wealthier ones are more likely to be represented in the big data sample, compared with the poorer ones. We compared expenditures of households in our big data sample based on expenditures from the VAT system as a share of their expenditures based on the 2018 HSES data and we found that on average, the ratio of VAT-based expenditure to HSES-based expenditure was 96.2 percent for female-headed households, whereas it was only 64.6 percent for male-headed households. Thus, female-headed households register their expenditures in the VAT system more fully compared with that of male-headed households.

For these reasons, the changes in spending of male- versus female-headed households are more meaningful compared with the absolute levels of spending. During 2020, we find that spending of female-headed households changed mostly in tandem of the spending of male-headed households.

²⁹ However, variations from month to month are visible in the figures in this section.

Figure 5. Mean expenditure by household heads' gender in 2019-2020, with confidence intervals



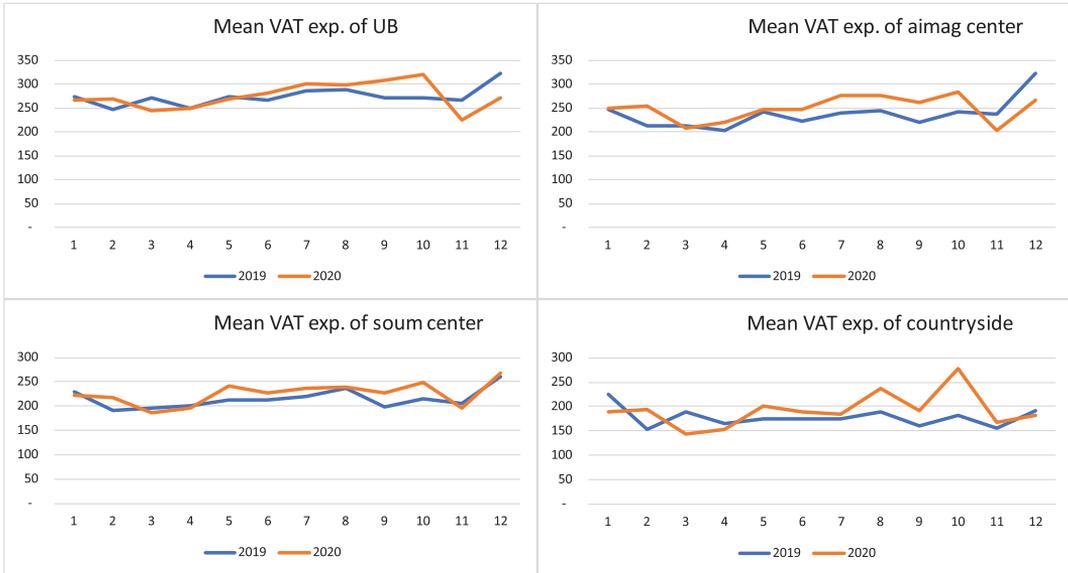
Source: Authors' estimates using the VAT data sample

4.1.2. Expenditure by rural-urban areas

Expenditure by household location shows that spending by households in urban areas - the capital city of Ulaanbaatar and aimag centers - followed a similar pattern, although the levels of spending in aimag centers are lower (See Figure 6). Rural households' expenditure - soum and countryside³⁰ - is lower than that of Ulaanbaatar and aimag centers, and follow a similar pattern with each other, showing that the soum centres' economy is closely tied to livestock herding. Rural households' expenditure shows high variability, which may be explained by two factors: rural nomadic households "accumulate" their spending needs, and once they visit soum centres, they make purchases in bulk; and the number of rural households in the big data sample is small, which also contributes to a less smooth pattern of consumption. Another important observation is that rural households' expenditure declined in March and April, at the time when the pandemic was raging in China, and when strict border closures were instituted in Mongolia. This means that herders have become unable to sell their produce, such as cashmere, wool, and hides, which significantly affected their income and consumption. Following a recovery in consumption in later months, soum and countryside households' consumption declined again in November and December, with a month' lag following the decline in the capital city and aimag centres' consumption, although the decline was not as pronounced as in urban areas. The main reason for rural consumption being affected less was that the lockdowns were predominantly instituted in Ulaanbaatar and a few aimag centres.

³⁰ Countryside households are nomadic herder households.

Figure 6. Mean expenditure by household location



Source: Authors' estimates using the VAT data sample. The sample is 4,463 households

Table 4. Changes in expenditure during the second lockdown in 2020

Mean monthly expenditure per person

	UB	Aimag center	Soum center	Countryside
November	-15.5%	-13.8%	-4.3%	7%
December	-15.7%	-17.6%	3%	-4.6%

Source: Authors' estimates using the VAT data sample. The sample is 4,463 households

For instance, during November and December of 2020, spending by households in the capital city of Ulaanbaatar fell by 15.5 and 15.7 percent, respectively.³¹ For households in aimag centers the decline was 13.8 and 17.6 percent respectively.

Changes in median spending are similar to the changes in mean spending. However, when we look at median expenditure, rural households' expenditure moves in a similar fashion with that of households in soum centres.

4.1.3. Expenditure by the number of children in the household

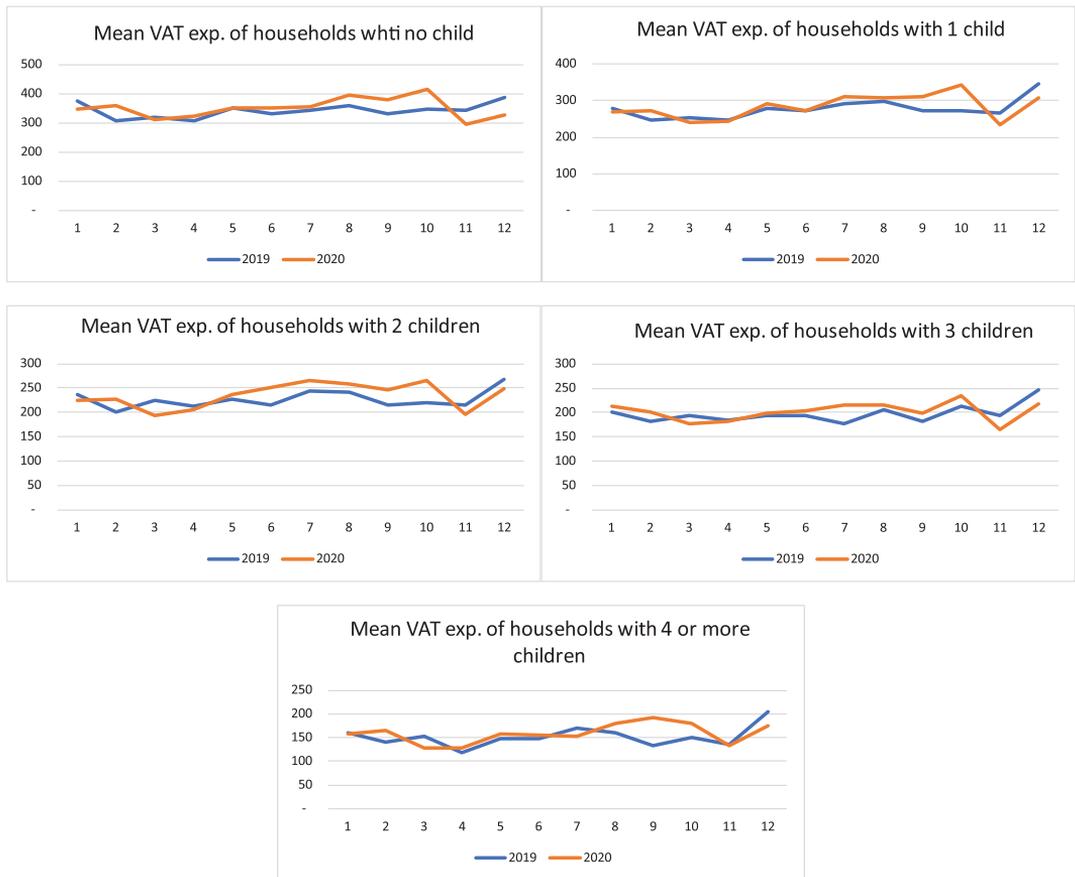
Spending of households by the number of children shows that the larger is the number of children in the household, the lower is per-person expenditure, with childless households having the highest per-person expenditure. Other observations can also be made. For instance, per-person spending of households with one child is between 250-300 thousand MNT per month, whereas it hovers between 100-200

³¹ per person, mean expenditure

thousand MNT per month in households with four or more children. This means that a child money benefit of 100 thousand MNT per month is sizeable, accounting for over 25 percent; for some families, it accounts for up to 100 percent of expenditure. For households with 1-2 children, the child benefit is unlikely to have made much difference, as their pattern of changes in expenditure has been more or less similar to that of childless households. Nevertheless, even after the significant increase earlier in the year, the child allowances and other social benefits were not sufficient to protect incomes during the second, stricter lockdown - as expenditure of most households - with and without children - fell sharply in November and December, by 7-16 percent year-on-year.

Figure 7. Expenditure of households by the number of children

Mean monthly expenditure per person



Source: Authors' estimates

4.1.4. Expenditure by quintiles

We also analyzed spending of households by quintiles, determined based on HSES-2018³², estimating year-on-year changes in their monthly spending in 2020 using big data (See Table 5). First, the two lockdowns had an immediate, large impact on the spending of households across all quintiles (with February being moderated due to the Lunar New Year). The first lockdown (January-March) affected much more the households in the bottom two quintiles, whereas during the second lockdown (November-December), their spending had still declined significantly (by 10 percent in November and 8 percent in December), but by less than that of the third and fourth quintiles. Considering that overall, the second lockdown affected spending by more than the first, it could be that the poorest households would have declined by even more if it was not for the increase in social assistance.

Second, starting in April, when the lockdown was relaxed and the increase in social assistance was announced, spending started to rise. But for the poorest households, the increase in spending did not occur until May – which is when the increased benefits were actually distributed - and for the second poorest quintile, until June. So it can be deduced that spending of the poorest households responds directly to the distribution of social benefits – and this spending increased by 8-20 percent year-on-year during the spring, summer, and fall months of 2020.

Third, households in the Q2 quantile reduced their expenditure more heavily than households in Q1. Their spending was higher than the pre-COVID period only in September and October 2020. Thus, the spending of the second poorest quintile was not so responsive to the distribution of social benefits, increasing only with a considerable time lag and not by as much as spending of the poorest households. For instance, whereas in July-October, spending further increased, the second poorest quintile's spending did not increase until September.

Third, spending of the third and fourth quintiles followed a similar pattern with each other, although the magnitudes were different. Monthly expenditures of households in the middle quintiles, Q3 and Q4, rose year-on-year during most of the year, except in March and during the second lockdown period. In some months, their spending increased by 32 and 38 percent year-on-year.

Fourth, households in the richest quintile, Q5, reduced their spending in January, March, May, and also during the second lockdown months. Even though they increased spending in the summer and fall, they did so by less than households in the other quantiles. Thus, spending of the richest quintile followed a very different pattern – it was reduced during both lockdowns to a greater extent and rose only modestly in the middle of the year.

Lastly, all quintiles had substantially reduced their spending during the second lockdown by 8-15 percent – in the presence of increased social benefits, which

³² In other words, the households were allocated into quintiles based on their consumption according to the HSES-2018 data, not the VAT data.

means that social benefits were no longer enough to cancel the effect of the lockdowns.

Table 5. Changes in mean household spending by quintile in 2020 compared with 2019

2020 months	Q1	Q2	Q3	Q4	Q5
1	-7%	-9%	10%	0%	-10%
2	18%	16%	14%	11%	10%
3	-10%	-15%	-4%	-2%	-11%
4	-1%	-6%	2%	4%	1%
5	8%	0%	11%	3%	-1%
6	20%	3%	13%	5%	1%
7	3%	-4%	8%	12%	15%
8	16%	-1%	6%	13%	3%
9	9%	15%	32%	17%	6%
10	18%	24%	15%	38%	8%
11	-10%	-8%	-11%	-10%	-15%
12	-8%	-13%	-15%	-11%	-9%

Source: Authors' estimates using the VAT data sample. The sample is 4,463 households.

Our findings are corroborated with findings of a study by the NSO and the World Bank - a joint COVID-19 Household Response Phone Survey (HRPS). The 1st round was conducted in May 2020 with 1,333 households and the second round in September 2020 with 1,212 households. The 1st round found that since the beginning of the COVID-19 pandemic compared to the same period of the year prior, 16 percent of households had no income, 73 percent of their business revenue in severe decline; moreover, 70 percent of herder families' income was in steep decline (NSO, World Bank, July 2020). According to the findings of the 2nd round of the research, about 95 percent of wage workers were able to work normally, 76 percent of households were able to maintain the same level of wage income since June – whereas only 58 percent did in Round 1 after the pandemic (NSO, World Bank, December 2020).

4.1.5. Spending versus income: the savings and loans factor

In poverty and inequality analysis, it is assumed that consumption is a proxy for income, and thus income poverty and inequality are derived based on consumption data. One of the major differences between income and consumption comes from savings/ investment and loans.

$$Income = Consumption + Increase in saving/ investment - Increase in borrowing$$

Our big data sample does not provide information on savings, investment, and loans. The national data on the balance sheets of the banking system provides one

possibility of gauging the changes in saving and borrowing of households. It shows that in 2020, the balance of loans by individuals was reduced by 969.7 billion MNT (\$344 million), that of non-performing loans increased by 234.9 billion MNT (\$83.4 million), while the balance of savings increased by 3.56 trillion MNT (\$1,263.1 million) (Mongolbank, 2020a). Thus, in 2020, households increased savings and repaid loans, which indicates that for the economy as a whole, household incomes have likely increased by more than consumption. Especially for the richest quintile, the reduction in consumption during lockdowns has likely been the outcome of these households reducing consumption, but increasing savings. At the same time, the amount of poor-quality loans also increased in 2020, indicating financial distress, which has likely affected households in poorer or less wealthy quintiles.

Although the banking system loans and savings are not the only types of loans and savings made by households, in Mongolia, they likely account for a large share of overall loans and savings in the economy. The extent of banking coverage is quite significant in Mongolia, although the poorest are less likely to be served by the banking system. About 93 percent of the adult population of Mongolia held a bank account (ADB, 2019). Moreover, the coverage by online banking is high and has grown rapidly - as of September 2020, there were 2.5 million internet banking users and 1.3 million mobile banking users at commercial banks providing mobile banking (Mongolbank, 2020b).³³

4.2. Findings on poverty

Our third research question was to estimate poverty and inequality in Mongolia using big data and to examine changes in them over the COVID-19 period.

We estimated the poverty headcount rates by using the official national poverty line, which was equal to 166,580.3 MNT per capita per month at 2018 prices.³⁴

Real Gross Domestic Product (GDP) increased by 5.2 percent in 2019, while the consumer price index (CPI) in 2019 was 5.2 percent higher than in 2018 (NSO, 2020). Moreover, while some sectors did better than others, especially in some quarters of the year, differences are not vast. Considering these changes and the low growth elasticity of poverty estimates observed in Mongolia (Uochi 2020, page 25), researchers considered the HSES data for 2018 as a reasonable representation of the situation before the shock caused by the pandemic (ADB, 2020).

Using the poverty line as a benchmark, we estimated the poverty headcount rates in Mongolia with our big data sample, which provides nearly real-time expenditure data.³⁵

³³ Mongolia's adult population (aged 20 years and older) in 2019 was 2.0 million. National Statistical Office Mongolia, available at 1212.mn. Last accessed on 8 March 2021.

³⁴ Uochi, I. (2020): Mongolia Poverty Update 2018 (English). Washington D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/532121589213323583/Mongolia-Poverty-Update-2018>.

³⁵ Nevertheless, this data was given to NUM researchers aggregated into months, because consumers tend to scan their E-receipts at the end of the month when rebates and lotteries are announced.

First, there is a significant difference in poverty headcount rate levels estimated using household surveys and big data. The official poverty rate based on the HSES-2018 per person consumption data is 28.4 percent. Big data estimate of the poverty rate in 2018 was estimated at 56.6 percent or 56.6 percent of the total individuals lives below the poverty line of 166,580.3 MNT. However, such a big difference in the poverty rate is not a cause for concern. There are several reasons for this.

The first reason is that the HSES-2018 poverty rate is estimated using households' all expenditure, regardless of whether they register their purchase to the E-receipts system. The big data estimate of poverty is based only on expenditures registered by consumers. Therefore, much of the consumption may not be registered in the E-receipts system. This is especially the case for low-income households who purchase groceries from the food markets with informal vendors.

The second reason is that HSE surveys also account for their own production and consumption. Rural herder households use a large share of food, meat, milk products from their own production, thus their expenditure per person is lower than urban family consumption. For instance, nearly 80 percent of the meat consumption of herder households comes from their own livestock production. In contrast, urban residents who receive cash income from wage or business activities pay almost every food item for themselves.

The third reason is that the HSE surveys define consumption (expenditure) more broadly than the VAT system. For instance, the HSES-2018 accounts for expenditure on housing, while the VAT system does not register some of these expenditures.

Thus, especially poorer households are not likely to register their expenditures on purchased food, while self-produced food is excluded altogether from the VAT system, their spending derived from the VAT system is bound to be lower than that from the HSES. In addition, across the board, households do not register some expenditure on durable goods and housing because of the way VAT law and system are designed. Therefore, this further reduces the consumption based on the VAT big data sample, relative to the HSES.

All three factors above – under-reporting of expenditures in the VAT system, exclusion of whole categories of expenditures, and exclusion of self-production and consumption – means that poverty rate calculated using the same official poverty line as is used in HSES will necessarily be higher compared with the HSES-based poverty rate. This does not mean that poverty is indeed higher and that VAT-based data is better at detecting poverty than HSES. It simply means that the two datasets compare different types of consumption.

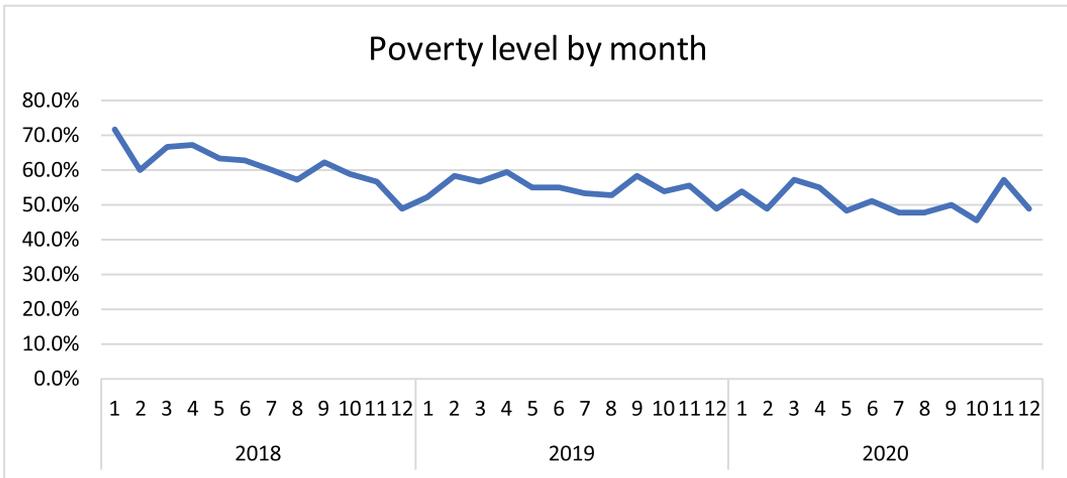
Therefore, it is more meaningful to examine changes in poverty over time, rather than the level of poverty. We do so using the same, consistent data set – in this case, the VAT-based big data sample. Using this data set, we estimate that the poverty rate was 56.6 percent in 2018, 49.4 percent in 2019, and 44.6 percent in 2020. In other words, the poverty rate, estimated based on the VAT big data sample, has

somewhat declined in 2020, the COVID-19 year, even though the rate of decline in poverty was less in 2020 than in 2019.

Although the finding that the poverty rate declined in 2020 in Mongolia is counterintuitive, it is nevertheless plausible because of sizeable transfers made by the government to households in 2020 as part of the economic stimulus.

Using this data set, we were also able to estimate monthly poverty rates. During 2020, the COVID-19 pandemic year, the poverty rate was 53.9 percent in January 2020; it declined through the year reaching levels lower than in 2019 and 2018 (See Figure 8). But during the second, stricter lockdown in November 2020, poverty increased to 57.6 percent before declining again to 49.4 percent in December.

Figure 8. The poverty headcount rate, by month



Source: Authors' estimates for the VAT data sample

We also estimated poverty rates from the VAT big data sample by location. As shown in Table 6, poverty in urban areas is lower than in rural areas. Geographically, the more rural is the location, the higher is the incidence of poverty. In Ulaanbaatar, the poverty rate is the lowest, followed by that in aimag centres, and then soum centres; in the countryside, the poverty rate is the highest. While these patterns are consistent with patterns of poverty from the HSES-2018, when interpreting these findings, it should be borne in mind that the share of food consumption that is not registered in the VAT system is likely to also be the highest among countryside (herder) households due to own production and consumption. Thus, the big data set underestimates the consumption of herder households (countryside) and overestimates their poverty.

Table 6. Poverty rate by location, based on expenditure

	2018	2019	2020
Urban	49.9%	44.3%	40.3%
Rural	68.9%	58.9%	52.6%
Ulaanbaatar	45.9%	41.3%	38.6%
Aimag center	58.5%	50.7%	44.0%
Soum center	67.1%	56.5%	51.9%
Countryside	73.0%	64.2%	54.1%

Source: Authors' estimates based on the VAT big data sample

4.3. Findings on inequality

The third and fourth research questions were to estimate inequality and its changes using big data, and to examine if there are differences in estimations using household surveys and big data.

4.3.1. Measures of inequality using indices

To answer these research questions, we measured the Gini, Theil, and Palma indices. The yearly inequality indices are shown in Table 7 and the monthly inequality indices are shown in Figure 9. All three indices of inequality using the big data sample are higher than the official estimates based on the HSES-2018 data. Similar to the differences in poverty estimates, it is not surprising that there is a big difference in inequality estimates using big data and HSES-2018. The estimated Gini coefficient for the big data sample is 0.474 in 2018, whereas for the HSES-2018, per person expenditure estimate, it is 0.352 - thus the Gini coefficient estimated for 2018 expenditure big data is higher by 0.12 percentage points than the estimate of expenditure inequality based on HSES-2018 data.³⁶

The Theil index gives a measure of the dispersion of expenditures between individuals or households. When the distribution of income is equal, each person or household has the same share of the overall available income or expenditure. The fact that the gap between the Theil index calculated based on the VAT data in 2018, and that based on the HSES-2018 is so large (0.410 versus 0.215) may be due to wealthier households being better represented in the VAT big data sample, combined with large variability in registration of their purchases in the VAT system by these households.

³⁶ Estimates of poverty and inequality are usually done on the basis of consumption, which is different from expenditure. In our case, since the big data sample provides data only on expenditure, not on total consumption, the directly comparable measure with HSES-2018 is the one using expenditure. Consumption-based inequality indices in the HSES-2018 data set were: Gini coefficient 0.327, Theil index 0.192, and Palma ratio 2.45

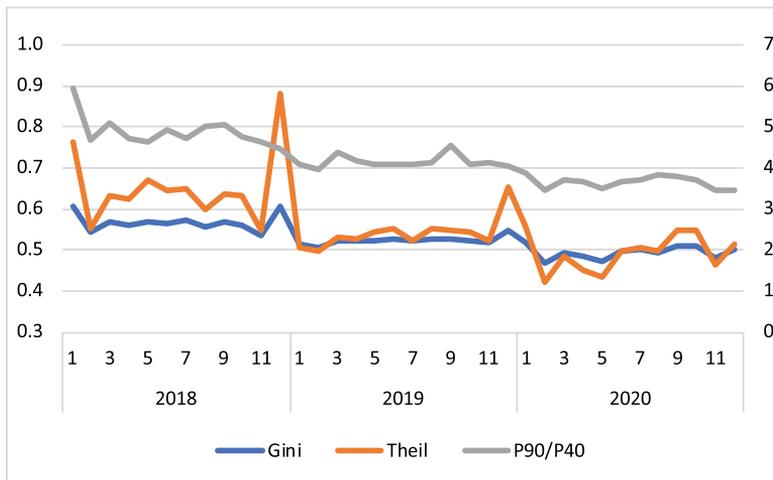
Table 7. Yearly indices of inequality

Big data	Observation	Gini	Theil (GE1)	Palma (p90/P40)
2018 expenditure per person	4463	0.474	0.410	3.86
2019 expenditure per person	4463	0.445	0.362	3.47
2020 expenditure per person	4463	0.419	0.329	3.15

2018 HSES data	Observation	Gini	Theil (GE1)	Palma (p90/P40)
HH expenditure per person	16,454	0.352	0.215	2.64

Source: Authors' estimates from the VAT big data sample. For the HSES-2018, the estimates were taken from NSO, 2019.

Figure 9. Monthly indices of inequality in expenditure



Source: Authors' estimates for the big data sample.

Underreporting of spending by poorer households will result in overestimation of inequality using the big data sample. As noted in the previous section, poor urban households buying food from informal market stalls, as well as herder households consuming their own food production (who tend to be poorer than the general population) means that their consumption is underreported in the VAT big data sample, thus overestimating inequality relative to the HSE surveys.

At the same time, underreporting of spending by richer households will result in underestimation of inequality estimates using the big data. Wealthier households are more likely not to register their purchases in the E-receipt system at all or partially, because the VAT rebate or lottery may not be a sufficient incentive for them to do so, resulting in underreporting. We did not find ways to detect such households and their rate of registration of purchases. Another factor is that privately provided health and education spending, which is more likely to be used by wealthier households, is exempt from VAT. Therefore, this results in underestimation of their spending and, consequently, underestimation of inequality.

Thus, due to VAT-based expenditure data differing from HSES-based consumption data, inequality measures using these two data sets will necessarily differ. It is more meaningful, therefore, to examine the trends in inequality over time using the VAT big data sample.

In 2019 and 2020, the Gini index declined slightly to 0.445 and 0.419, respectively. In other words, during the COVID-19 period, inequality in expenditure among households in Mongolia declined compared with 2018 and 2019.

Big data estimates based on per person expenditures suggest that in 2018, the Theil index was 0.410 and in 2020 the index declined to 0.329. Some researchers might infer that in Mongolia, during the COVID-19 period, inequality in individuals' spending has declined compared to the pre-COVID-19 period. As shown in Table 5, during COVID-19 months, spending of the poorest quantile declined by less compared with households' in Q2 and households in Q5.

The Palma ratio was 3.86 in 2018. In other words, the richest 10 percent of individuals in the big data sample spend 3.8 times more than the poorest 40 percent of persons in the sample. In 2020, the index declined to 3.15.

Thus, changes in all three indices show that inequality of expenditures in Mongolia declined in 2020.

The decline in inequality of expenditure can be attributed to the mitigation impact of the Government's pro-poor policies – generous social assistance as part of the fiscal stimulus package, as was also corroborated in recent research (ADB, 2020). Indeed, we have seen from See Table 5 that after receiving the increased amount of welfare assistance from the Government, the poorest two quintiles increased their expenditures, and during the second lockdown at the end of 2020, their expenditures, in general, declined less than that of the three wealthier quintiles. In addition, it can be explained by the decline in the monthly expenditure by more affluent households in 2020 (See Table 5, Q5 column). Lockdowns and border closures mean that the opportunities for wealthier households to spend became severely limited.

Also, due to the government progressively increasing the coverage of VAT from 2018, more goods and services have become taxable under VAT, prompting households – particularly poorer households - to register more of their expenditures in the VAT

system, which may have reduced the inequality of measured expenditures.

Finally, the decline in inequality as measured by the three indices above only shows inequality in expenditure. Inequality in income may have increased. Since wealthier households have more room for saving, it could be that they saved more during 2020, and thus the income inequality, which cannot be observed from the big data (expenditure data), could have increased. Indeed, it is likely that in a time of extreme uncertainty, such as that presented by the pandemic, those who could afford could have made precautionary savings. Considering that wealthier households are likely to have incomes that are higher than expenditures, and that the overall net savings in the economy grew in 2020, we can deduce that it is the wealthier households that are more likely to have saved, resulting in underestimation of inequality.

Put together, this could mean that inequality of income could have risen in 2020, even though inequality in expenditure, as measured by VAT data, has declined.

4.3.2. Measures of inequality using distribution

We also measure inequality using measures of distribution – by quintiles and by density distribution. Moreover, the big data sample provides a rare chance to construct a panel data of households, from which we can see how inequality has changed over time.

The big data sample is derived from (is a subset of) the HSES-2018, and since households in the HSES-2018 sample are categorized into quintiles by their consumption, we use these HSES-based quintiles as the basis. In addition, the big data also includes expenditures of households in 2018-2020, which allows us to also categorize them into a different set of quintiles based on the VAT-based expenditures of these respective years.

Table 8 shows how quintiles based on the 2018 VAT big data match against the quintiles based on HSES-2018 consumption. The percentages of matched households – those in the same consumption, expenditure quintiles in both samples – are shown colored in grey in the diagonal area.

For example, out of all households in the poorest quintile according to the HSES-2018 consumption, 42.2 percent remain in the poorest quintile according to the VAT big data, while 20.5 percent appear in the second poorest quintile. The remaining 37.4 percent show up in the third, fourth, and the richest quintiles of the VAT big data sample. This means that many of the poorest households had higher expenditures registered in the E-receipts system compared with those in the higher quintiles – or it could also mean that households in the richer quintiles did not register many of their expenditures, so by VAT-measured expenditure, they appear poorer. This could also arise, for example, if many of the expenditures of the richer quintiles are expenditures on assets, such as housing or vehicles, as well as education or health spending, which are not covered by the VAT big data adequately or at all.

Similarly, 50.4 percent of households in the richest quintile per the HSES-2018 consumption remain in the richest quintile according to the VAT, while 23.7 percent of them appear in the second richest quintile according to the VAT, with the rest distributed between the three poorer quintiles. Thus, this table shows that there are substantial differences between HSES and big data-based data – due to some categories of consumption not being captured in big data (VAT) system, or substantial under- or over-reporting. Therefore, the VAT data cannot be used to accurately measure inequality – and thus cannot substitute the HSE survey data.

Table 8. Matching of 2018 HSES consumption quintiles and big data expenditure quintiles

		2018 big data expenditure-based quintiles					
		Q1	Q2	Q3	Q4	Q5	Total
HSES consumption- based quintiles	Q1	42.2%	20.5%	23.6%	9.7%	4.1%	100.0%
	Q2	18.1%	29.1%	23.4%	18.6%	10.8%	100.0%
	Q3	16.1%	20.5%	23.4%	25.0%	15.0%	100.0%
	Q4	8.7%	17.3%	21.1%	26.1%	26.8%	100.0%
	Q5	3.3%	8.0%	14.6%	23.7%	50.4%	100.0%
TOTAL		13.9%	18.3%	21.7%	23.1%	23.1%	100.0%

Note: Based on mean expenditures per quintile

Source: Authors' estimates from the HSES-2018 and the VAT data sample

Therefore, it is more meaningful to look at movements between quintiles using the consistent data source - using the panel, or longitudinal, data on expenditure in 2018-2020, generated from the big data sample. Using this rare data, we looked at how many households have moved between expenditure quintiles between 2018 and 2020, based only on the expenditures based on VAT (See Table 9). The table shows that there has been a considerable movement between quintiles. About 60 percent of the households that were in the poorest expenditure quintile (Q1) in 2018 stayed in the same quintile in 2020. In other words, they did not move upward in terms of expenditure and consumption in 2020. However, 22.7 percent of them moved to the second quintile (Q2), 10 percent to the middle quintile, and a small share, or 4.8 and 2.8 percent, respectively, moved into upper quintiles (Q4 and Q5). Similarly, just over 60 percent of the richest quintile households (by expenditure) in 2018 stayed in the same quintile in 2020, 24.2 percent were in the Q4 quintile, and the remaining share shows up in the poorer quintiles in 2020. These movements conceal both genuine changes in expenditure, indicating social mobility, as well as increased or reduced registration of expenditures by households in the E-receipts system.

Table 9. Movement of households between expenditure quintiles between 2018 and 2020

		2020 Big data expenditure-based quintiles					Total
		Q1	Q2	Q3	Q4	Q5	
2018 big data expenditure quintiles	Q1	59.8%	22.7%	10.0%	4.8%	2.8%	100.0%
	Q2	24.0%	33.5%	26.6%	11.9%	4.0%	100.0%
	Q3	11.0%	27.1%	30.2%	22.9%	8.8%	100.0%
	Q4	4.0%	12.8%	23.8%	36.3%	23.2%	100.0%
	Q5	2.0%	3.0%	9.4%	24.2%	61.3%	100.0%

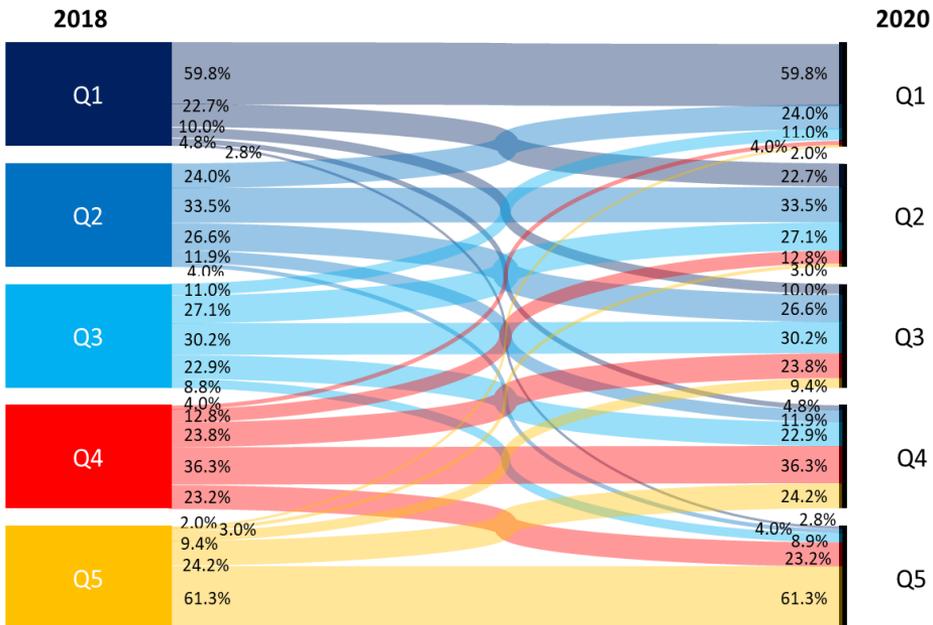
Note: Based on mean expenditures per quintile

Source: Authors' estimates from the VAT data sample

How about households in the middle, in both samples? How have they changed in terms of spending? Did they move upwards or downwards in 2020? 1/3 of households in Q3 in 2018 stayed in the same expenditure level in 2020, but about 23 percent and 9 percent moved into the upper quintiles (Q4 and Q5). Interestingly, 61.3 percent of households in Q5 in 2018 stayed in the same highest expenditure level in 2020, but about 2 and 3 percent of them moved down into lower quintiles (Q1 and Q2).

Figure 10 shows the upward and downward movement of households in the country between quintiles between 2018 and 2020.

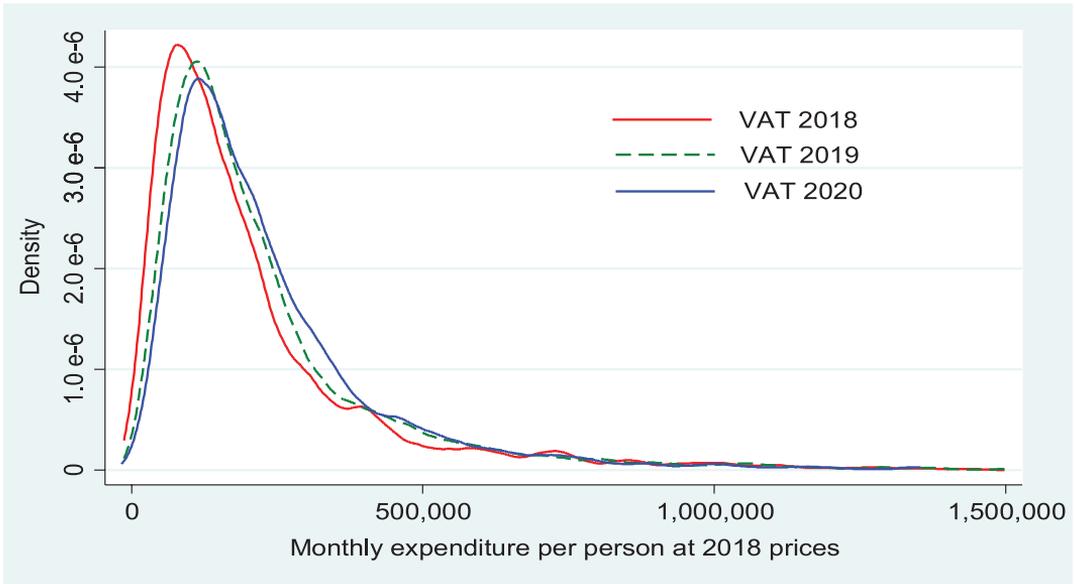
Figure 10. Movement of households expenditure quintiles between 2018 and 2020



Note: Based on mean expenditures per quintile Source: Authors' estimates from the big data sample

A more granular view of the changes in inequality can also be seen from the shift in the expenditure density curves from 2018 to 2020. The density curve moved to the right and became slightly more spread out in 2020 compared with 2018, indicating that the expenditure levels increased, and their distribution became slightly more equal.

Figure 11. Expenditure density curves in 2018-2020



SECTION 5. CONCLUSIONS

5.1. Conclusions on research findings

In 2020, the world experienced an unprecedented pandemic and an economic shock of an enormous magnitude. In this report, we examined the impact of COVID-19 on consumption, poverty, and inequality in Mongolia throughout 2020 using big data. The existence of the VAT E-receipts system in Mongolia which covers a large and growing number of consumers, provided a unique opportunity to estimate consumption (expenditure), as well as poverty and inequality during the pandemic using actual data, as opposed to modeled estimates. The big data provides using a consistent panel (longitudinal) data set, which enables measuring these indicators with monthly frequency.

We found that during most months of 2020, expenditure increased, while poverty and inequality declined, compared with the same months of 2019, based on the VAT data. Big data revealed that total spending among households fell sharply in January

2020, but nearly restored back to pre-COVID levels by late June 2020. These findings are consistent with empirical evidence from the recent research carried out in Mongolia (ADB, 2020, UNFPA, 2020). However, the two strict lockdown periods in the beginning and the end of 2020 had a large and immediate impact, reducing consumption and increasing poverty and inequality. There was also enormous volatility in consumption during 2020 from one month to another. A more nuanced explanation is that until November 2020, Mongolia has instituted strict border control and quarantine measures from as early as February 2020, which allowed the country to open movement and economic activities after the initial lockdown of January-March. Therefore, while economic activities dependent on foreign trade suffered, especially early in the year, the domestic-oriented economic activities were carried out as usual for most of 2020.

Overall, when taken on an annual basis, expenditure was higher in 2020 than in 2019, while poverty and inequality were lower because the government has taken large size of stimulus measures; and some economic activities were less disrupted until domestic transmission in November 2020.

Our findings also show that the fiscal stimulus by the government, specifically, spending on social assistance has been effective – it had an immediate and sizeable effect on increasing consumption and protecting it from falling, especially for the poorest, and may have also contributed to the increase in consumption of the non-poor. However, social assistance measures were not enough to outweigh the negative impact of strict lockdowns on consumption.

Although the levels of poverty and inequality measured using the VAT big data sample were much higher compared with those based on HSES-2018, this was expected, because the VAT data sample does not allow measuring all types of expenditure that are measured using statistical household surveys. More important than the differences in levels between VAT data-based and official survey-based estimates of poverty and inequality are the changes in poverty and inequality in expenditure, which could be measured on a monthly basis using the big data sample.

Using the VAT big data sample, we found that the poverty rate in 2018 was 56.6 percent, which declined to 49.4 percent in 2019 and further declined to 44.6 percent in 2020. Thus, despite the year 2020 bringing a large economic shock to Mongolia, poverty declined somewhat in 2020, which can be attributed in a large part to social protection measures taken by the government, as well as preventive public health measures that were taken early on.

Inequality – measured using the Gini coefficient, as well as Theil and Palma indices – has also somewhat declined in 2020. For instance, the Gini coefficient measured using the big data sample (which is significantly higher than that using the household survey data) was 0.474 in 2018, 0.445 in 2019, and 0.419 in 2020.

There are **several caveats** to accompany these findings. In our case, the VAT big

data measures only expenditure and does not fully account for consumption, even though the coverage of the VAT system is high and is rapidly growing. It can also conceal systematic overreporting and underreporting of consumption – and to the extent to which such behaviour is exhibited by poorer or richer people – can result in distorted measures of poverty and inequality. ***Thus, VAT data-based analysis cannot substitute for household surveys conducted using rigorous statistical methodologies.*** Moreover, inequality and poverty using VAT data are based on expenditure, not income; therefore, while inequality of expenditure may have declined, the inequality of income might have increased.

Nevertheless, it provides an important complementary source of data to gauge consumption, poverty, and inequality of households, at a much higher frequency than was possible until now. Significantly, with big data - such as the VAT data set in Mongolia - poverty and inequality are no longer slow-moving indicators. This study demonstrated that their changes can be measured as frequently as on a monthly basis. In this regard, it allows examining how do poverty and inequality respond to policy changes, such as public spending on social protection, or tax-related measures. In this regard, the big data-based analysis, when implemented in a rigorous manner, can become an important tool to support public policy-making.

5.2. Recommendations on big data use

The VAT big data can provide valuable information for policymakers. At the same time, it is of utmost importance to ensure respect for the privacy and confidentiality of individual data. Through the experience of conducting this study, the first of its kind in Mongolia, we designed and implemented protocols to ensure the protection of privacy and confidentiality, while conducting analysis that can help inform policy decisions.

Such protocols can be used if similar research is to be done in the future (See Section 2.2). Below, we also identify ways in which data collection, processing, and sharing processes can be improved further.

First, the rollout of POS machines should be continued to ensure that wider categories of spending can be registered. If the cost of investing in a POS machine is a deterrent, POS-sharing systems can be instituted between small and informal enterprises, for example, at markets.

Second, given that poorer and rural households have less access to the internet and smartphones, which prevents their use of the E-receipts system, the government can consider providing other means for users without smartphones to register their purchases. This can be done, for example, by improving the web-based E-receipts interface (in addition to the smartphone app), providing access to the E-receipts system from public service automated kiosks used increasingly in Mongolia and, more generally, expanding access to the internet for all.

Third, better ways should be devised to enable categorization of spending by type,

while at the same time, not burdening consumers with excessive steps needed to complete the registrations.

Fourth, the government should consider enabling anonymized data from both consumers and companies (buyers' and sellers' side) available for research, since the companies' records are much more complete compared with consumers' records. The government can consider making such anonymized data publicly available, similar to HSE surveys microdata.

Fifth, data processing capacities at the Tax Data Center should be improved in terms of hardware, software, and staffing. Currently, retrieval of sample data requires stopping the system which runs 24/7; instead, the system should enable retrieval of data while the system is running. More staffing capacities are needed to enable the processing of this valuable data.

Sixth, the methodologies used in this research can serve as the basis for establishing procedures and protocols to overcome technical and bureaucratic hurdles in analyzing and using data, while ensuring protection and confidentiality of big data from the VAT E-receipts system.

Effective collaboration between government agencies, researchers, and international organizations to make this data an important resource that will facilitate better and timely public policy decisions.

REFERENCES

- ADB (2019). Strengthening the Protection of Financial Consumers in Mongolia's Banking Sector. ADB East Asia Working Paper Series No 20. December 2019.
- ADB(2020). Building Capacity for an Effective Social Welfare System (Issue September). <https://www.adb.org/sites/default/files/project-documents/51387/51387-001-tacr-en.pdf>
- Bataa, Erdenebata and Batbekh Soyolmaa, (2020). Ковид-19 цар тахлын үеийн эдийн засгийн нөхцөл байдал, үзүүлэх нөлөө, хариу арга хэмжээний саналууд [Feedback and comments on response to the COVID affect to the economy], available from: https://www.researchgate.net/publication/347945492_Kovid-19_car_tahlyn_eijn_edijn_zasgijn_nhcl_bajdal_zleh_nl_hariu_arga_hemzeenij_sanaluud
- Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic opportunity in the United States: An intergenerational perspective. *The Quarterly Journal of Economics*, 135(2), 711-783.
- Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & TOI TEAM. (2020). The economic impacts of COVID-19: Evidence from a new public database built using private sector data. Working Paper 27431, National Bureau of Economic Research.
- Conceicao, P. N., & Ferreira, P. M. (2005). The Young Person's Guide to the Theil Index: Suggesting Intuitive Interpretations and Exploring Analytical Applications. *SSRN Electronic Journal*, January 2019. <https://doi.org/10.2139/ssrn.228703>
- Deaton, A., and S. Zaidi. 1999. Guidelines for Constructing Consumption Aggregates for Welfare Analysis. Princeton, Woodrow Wilson School - Development Studies.
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210). <https://doi.org/10.1126/science.1243089>
- Himelein, K. (2014). Weight Calculations for Panel Surveys with Subsampling and Split-off Tracking. *Statistics and Public Policy*, 1(1), 40–45. <https://doi.org/10.1080/2330443X.2013.856170>
- Mongolbank (2020b). Mobile Banking Report. 3rd quarter of 2020.
- Mongolbank (2020a). Consolidated Balance Sheets of the Banking System. December 2020. Available at: <https://www.mongolbank.mn/liststatistic.aspx?id=8>. Last accessed 8 March 2021.
- NSO (National Statistical Office of Mongolia). The World Bank, and UNDP (United Nations Development Programme). 2004. Main Report of Household Income and Expenditure Survey/Living Standards Measurement Survey 2002-2003. Ulaanbaatar: NSO.

NSO (National Statistical Office of Mongolia). 2019. Report of the Household Socio-Economic Survey 2018.

NSO (National Statistics Office of Mongolia). (2020a). Gross domestic product. Main Economic Indicators, 2020(4). <https://doi.org/10.1787/54ace363-en>

NSO (National Statistics Office of Mongolia). (2020b). Socio-Economic Situation of Mongolia, monthly bulletin (1-12), 2020. Available at: <https://www.1212.mn/BookLibrary.aspx?category=005&fbclid=IwAR3EQHutTrToQddOa4H1qKhkZ8Djz4Lb-VdN0eRvgdy5OiqOU6LItMUCBWg>

NSO (National Statistical Office of Mongolia). 2021. Retrieved from 1212.mn. Last accessed on 20 March 2021.

NSO and World Bank. Results of Mongolia COVID-19 Household Response Phone Survey (Round 1). Retrieved January 4, 2021, from <http://documents1.worldbank.org/curated/en/656061595316484647/pdf/Results-of-Mongolia-COVID-19-Household-Response-Phone-Survey-Round-1.pdf>

Filzmoser, P., Gussenbauer, J., & Templ, M. (2016). Detecting outliers in household consumption survey data. Vienna, Austria: Vienna University of Technology.

Uochi, I. (2020). Mongolia poverty update 2018 (No. 148387). <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/532121589213323583/mongolia-poverty-update-2018>

UNFPA. COVID-19 Situation Report No. 6 (2020) <https://www.unfpa.org/resources/covid-19-situation-report-no-6-unfpa-mongolia>

ANNEXES

6.1. Methodology for processing big data

We combined two sources of data – VAT records housed at the Tax Data Center and the Household Socio-Economic Survey (HSES) 2018 data from the NSO.

The HSES-2018 was the latest household living standards survey conducted in Mongolia at the time when this research started. HSES-2018 is representative of the entire population and at the regional level through stratified random sampling.³⁷ The HSES-2018 sample included 16,454 households or 1.8 percent of all households in the country.

The HSES-2018 sample was used in our study as the validating data set. This was done for several reasons.

First, the Tax Data Center provided only a sample of the VAT data for this research – processing the full data set was not practical due to the size of the big data, risks involved in the transfer, and high requirements for computing capacity. Moreover, it was not necessary, because sampling from this big data set was sufficient for our research. Therefore, to sample the data from the VAT while being able to make inferences from this retrieved sample required applying a robust methodology of sampling. Since the HSES-2018 is representative of the overall population, a solution that we found was “pulling out” from the VAT data set the records for the HSES-2018 households. Doing this was equivalent to applying the same sampling methodology of HSES-2018 to the VAT data set.

The second reason to link the two data sets was that consumption, poverty, and inequality need to be calculated at the household level, whereas the VAT dataset contains data only on individuals. Therefore, linking of the data sets enabled to attribute individual-level data from the VAT data sample³⁸ to household-level data from the HSES. Such a method of linking datasets is described in Chetty and others (2014).

The rest of this section provides a step-by-step explanation of the process of matching big data from VAT with the data of the 2018 household survey.

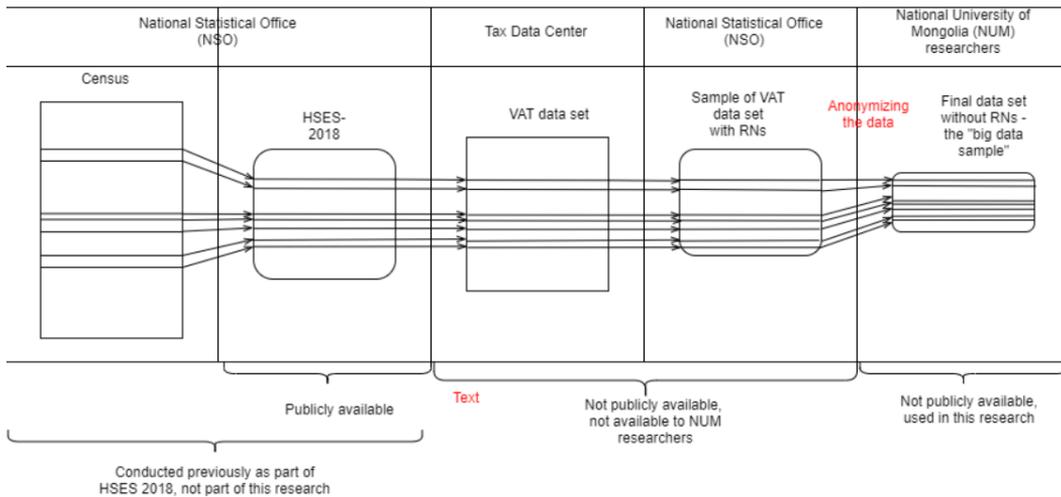
The data from the two sources were linked using a unique identifier called the Registration Number (RN), or the national ID number, that is assigned by the General Authority for State Registration to every newborn citizen of the country. The RN data is strictly confidential and is administered by the Population and Household Database (PHDB) of the NSO. To protect privacy and confidentiality of data and to adhere to the law to not to release RNs to unauthorized persons, protocols were designed and followed in this study. Only those members of the research team who had the authorization to access

³⁷ At the time when this study was started, in October 2020.

³⁸ When a person makes a purchase, the VAT receipts can be entered (scanned) into the E-receipts system by anyone in the household – the person himself/ herself, the spouse, the children or other family members. This is often done by women. However, in order for consumption or spending to be accurately measured, it needs to be first, aggregated for each household, and then divided by the number of people in the household.

the records with RNs as part of their official duties - the NSO and the Tax Data Centre officials - accessed the data with RNs and performed this data processing. After the sampling and first-stage processing were done and the data was anonymized, the data set was given to the researchers of the National University of Mongolia. The process of this sampling, processing, and anonymization is shown in Figure 12, and the steps are explained in detail in the rest of this section.

Figure 12. First-stage processing of the big data for this research



First, the NSO statisticians obtained authorization from the Population and Household Database (PHDB) of the NSO to obtain RNs of individuals in the HSES-2018 sample.³⁹ They retrieved the names and household numbers of 59,820 people in the HSES-2018 survey and sent them to the PHDB.

Second, out of 59,820 people, the PHDB data analysts identified only 42,991 individuals' RNs who belong to 13,733 households.⁴⁰ Thus, some data was lost (RNs could not be found) due to reasons such as misspelling of the first or last names.

Third, the NSO sent the RNs of these 42,991 individuals to the Tax Data Center for matching. The Tax Center's data scientists matched the RNs of these 42,991 individuals to the VAT database at the Tax Data Center, containing data on individuals who file E-receipts into the system, and retrieved data on 23,600 individuals along with their aggregated monthly spending.

Fourth, after receiving data on monthly spending from the E-receipts systems from the Tax Data Center, the NSO statisticians matched them with the data from the HSES-2018 sample, identifying that the 23,600 individuals belonged to 9,826 households. This data on individuals and households was anonymized (RNs and names removed) and provided to the NUM researchers.

³⁹ HSE Surveys do not collect information of RN of interviewers.

⁴⁰ This is referred to as the "matched HSES-2018 sample" of 13,733 households.

Fifth, the NUM researchers conducted second-stage processing of the data. They obtained publicly available HSES-2018 data containing data on individual and household characteristics such as gender, age, location, number of people and children in households, etc., thereby attributing anonymized spending data with data on these general characteristics.⁴¹ Further, to minimize the problem of having records on individuals with incomplete spending data,⁴² the NUM researchers filtered out individuals who do not register their receipts consistently. Out of the initial 9,826 households, there were only 7,127 households whose members used the E-receipt app at least once per year in consecutive three years 2018-2020. Filtering it even further, it was found that there are 5,069 households whose members used the E-receipts app at least once in a quarter, and 4,463 households whose members used the app at least once in a month during these three years. The final criteria were made stringent, leaving only those households which used the E-receipts app at least once a month. Thus, the final data set that we used in this analysis included 4,463 households, hereinafter referred to as “the big data sample” in this report (See Figure 12).

A detailed workflow of data sampling and the anonymizing process is shown below.

The resulting data set is a panel (longitudinal) data set of these households, containing data from both the HSES-2018 microdata, which is available publicly, as well as data on the total amount of expenditures from the VAT big data.

Table 10. First-stage processing of the VAT and HSES-2018 data

Step	Source of data	Recipient of data	Number of households	Number of individuals	Comments
1. Identify individuals in the HSES-2018 sample	NSO, from HSES 2018	NSO, PHDB	16,454	59,820	By matching HSES-2018 data of 59,820 individuals from 16,454 households backwards with the Census data, it was found that records of 42,991 individuals from 13,733 households had RNs. The rest could not be located due to name misspelling or other reasons. This step was performed by the NSO PHDB, authorized to have access to RNs.

⁴¹ The anonymized data of HSE surveys is publicly available in Mongolia.

⁴² People may choose to register only some of their purchases in the E-receipts system or may not register them consistently every month; therefore the E-receipts data does not provide complete data on individuals' spending.

COVID-19 AND HOUSEHOLDS' EXPENDITURE

2. Provide individual records with RNs to the Tax Data Center	NSO	Tax Data Center	13,733	42,991	The NSO PHDB then provided the RNs of the above 42,991 individuals to the Tax Data Center.
3. Matching data to draw a sample from the E-receipts database	Tax Data Center	NSO	9,826	23,600	Of the 42,991 individuals, the data on 23,600 individuals were found in the E-receipt database. This means the remaining 19,391 persons do not use E-receipts app.
4. Anonymize the big data sample data	NSO	National University of Mongolia researchers	9,826	23,600	This data set was anonymized (RNs removed) and provided by the NSO to the researchers.
5. Further narrow the data sample	NSO	National University of Mongolia researchers	9,826 ↓ 4,463	37,382 ⁴³ ↓ 17,607	To ensure the quality of the data set, the researchers further narrowed the data set to only those households that consistently register their receipts in the E-receipts system. It was found that in 2018-2020, there were: 7,127 households who used the E-receipt app at least once per year 5,069 households who used the E-receipt app at least once per quarter 4,463 households who used the E-receipt app at least once per month The final data set that was used in this research was that of 4,463 households (referred to in this report as the “big data set”).

⁴³ Whereas the VAT data set contained 9,826 households with 23,600 individuals, from HSES-2018 data set, it was determined that 37,382 individuals belonged to these 9,826 households. This is because the VAT data set contains data mostly on adults who make purchases. But the HSES data set contains data on all household members, including not only those who make purchases, but also children and elderly.

6.2. Second-stage data processing

The big data sample contained some outliers. For example, there were cases when a household had monthly expenditure of 100 million MNT,⁴⁴ whereas its usual monthly spending was about 1 million MNT. These outlying values can have a significant impact on analysis like inequality indicators or regression coefficients.

We used a univariate outlier detection method in which data values were transformed by the robust Box-Cox transformation (Filzmoser, Guesenbauer, and Templ, 2016).⁴⁵ For instance, we found that in 2016, there were 6 households whose E-receipts totalled more than 100 million MNT per month. Similarly, 12 households registered E-receipts worth more than 100 million MNT per month in 2017, 40 such households in 2018, 59 households in 2020. When we detected unusually high or low values, we tracked down the sectors in which the head of the household is employed from the HSES -2018 sample data and found that they mostly work in the service sector. Moreover, we investigated further to learn that in a large purchase-making household, not just their head of household, but also other members work in the service or sales industry. This means that service sector employees in business establishments may have (unlawfully) scanned other customers' E-receipts as their own. The Tax Data Center data scientists confirmed this assumption. Therefore, for such outliers, the expenditures were imputed to be in line with their usual expenditures.

6.3. Reweighting and adjustment for bias

The HSES-2018 data set of 16,454 households was a stratified random sample, representative at the national level, location level (Ulaanbatar, aimag, soum and countryside), as well as at the level of each aimag.

However, the big data sample (4,463 households, a subset of the HSES-2018) was biased because rural households, in particular, are less represented in the big data sample due to the lower rate of uptake of the VAT E-receipts system in rural areas. Among rural households, those living in the countryside (herder households) are especially poorly represented in our sample – countryside households in our sample represented only 8.3 percent of countryside households included in the HSES-2018 sample, a far lower percentage than the national average of 27.1 percent (See Table 11). Soum centre households are also less well represented than the national average, at 22.3 percent. In contrast, households from Ulaanbaatar and aimag centres are represented better than the national average, at 37.1 and 35.8 percent, respectively.

When such a biased sample is used, it cannot be representative of the whole population and inferences cannot be made of the findings. Therefore, the reweighting of households was done within the subsamples of each of the four

⁴⁴ 35,087 USD at the current exchange rate.

⁴⁵ We used the rule "median plus/minus 3 times interquartile range"

types of locations (Ulaanbaatar, aimag, soum and countryside levels).⁴⁶

For instance, household X in the HSES-2018 sample might have a weight of 50, indicating that it represents 50 households within the overall population in the country. However, household X has a certain probability of being further “selected” from the HSES-2018 sample – or the probability of ending up in the final data sample. For example, from Table 11, we can see that over 30 percent of households from HSES-2018 that live in Ulaanbaatar or aimag centres ended up in the final big data sample, whereas only 22.3 percent of households living in soum centres and 8.3 percent of countryside households ended up in the final big data sample. If our household X in the HSES-2018 survey that ended up in our big data sample was from Ulaanbaatar, it would represent about 135 households within the overall population ($50/0.371 \sim 135$); but if it was from the countryside, it would represent about 602 households ($50/0.083 \sim 602$).

These examples of reweighting are hypothetical and simplified.

Such reweighting was done by four types of location: Ulaanbaatar, aimag centre, soum centre and countryside. Subsequently, when estimating poverty and inequality, these household weights were taken into account, so that the findings are representative of the overall population and by location.

Table 2 shows the shares of households by location with and without the weights. When weighted, the VAT sample is very similar to the HSES sample, although differences still exist for soum and countryside households (See Appendix 1 for a more detailed description of weights of the big data sample).

Table 2. Distribution of households in the big data sample by location, before and after weighting

	Unweighted		Weighted	
	HSES	VAT	HSES	VAT
Ulaanbaatar	21.7%	29.7%	45.8%	44.4%
Aimag	32.8%	43.3%	20.5%	20.3%
Soum	25.4%	20.8%	18.1%	24.3%
Countryside	20.1%	6.1%	15.6%	11.0%
	100.0%	100.0%	100.0%	100.0%

6.4. Estimation methodologies

In this study, we sought to reproduce the methodology used by statistics offices to estimate consumption, poverty, and inequality, to the extent possible. Nevertheless, it should be noted that measures of consumption and, consequently, poverty and

⁴⁶ Although other groups might also be less represented, the reweighting was done for location only, following the methodology of weighting of the HSE surveys.

inequality estimated using the big data differ in important respects from those measures estimated through household surveys due to the underlying ways in which data is collected. This section details the methodology for estimating these measures, explaining the similarities and differences.

6.4.1. Estimation of consumption (expenditure)

Our measure of consumption generated from the VAT system is not equivalent to the measure of consumption estimated by HSE surveys. Major differences exist between consumption estimated using HSE surveys and that using the VAT data.

The VAT E-receipts system registers market transactions and thus does not include consumption of self-produced or owned goods, as well as “consumption” of own housing. Most herder households in Mongolia consume self-produced food – meat and dairy products, which normally account for a sizeable proportion of consumption (See Box 1). In household surveys conducted by statistical offices, consumption of self-produced food products or self-owned housing is estimated by imputation – by approximating to what would such households have spent on consumption of these goods if they did not self-produce or own them. However, in our big data sample, we do conduct such imputation – our estimates are based entirely on the big data – or the registration of expenditures by households in the VAT E-receipts database.

Roughly speaking, consumption of self-produced or owned goods constitutes a difference between consumption and expenditure. Whereas expenditure, or spending, represents consumption of only those goods purchased through market transactions, consumption constitutes the entirety of consumption of goods and services, including those obtained “outside” of the market. In other words

Consumption per HSES = Expenditure + Consumption of self-produced and owned goods

In this regard, the VAT big data sample provides only expenditures, which in our study is used as a proxy for consumption.

In addition, some types of transactions are not included in the VAT system because some goods and services are exempt from VAT and, therefore, consumers have no incentive to register them, as they receive no VAT refund. Some transactions are not included because they are carried out through small- and informal enterprises which do not have POS machines. Due to the government implementing drives to require small and informal retailers to install POS systems, an increasing number of such enterprises have started using them, but the coverage is always less than full because a portion of economic activities will always remain informal. For a more detailed overview of the types of goods and services covered by the VAT system in Mongolia, see Box 1.

One of our research questions was to examine how the categories of spending by consumers changed during COVID-19. Such tracking of categories of expenditures was attempted, but could not be conducted because the VAT data categories were limited (See Box 1).

Box 1. Types of consumption covered by the VAT system in Mongolia

There are two ways in which the VAT system enables the classification of expenditures into categories: through sellers and consumers.

The first type of classification is based on legal entities' registration type – depending on whether a company works in education, health, or trade sectors. However, if a firm trades many different types of goods, the E-receipt is classified into the category of “others”. For instance, supermarkets and department stores that account for a large share of purchases, are classified as “other” and therefore, purchases made at these stores are classified as “other”. However, as a result, the VAT data categorization becomes meaningless, with about nearly 90 percent of expenditures classified into the “other” category.

The second type of classification is based on consumers themselves categorizing their spending while scanning their receipts. This feature of the VAT E-receipts system was introduced in January 2019. After scanning the receipts into the E-receipts app, consumers now have an option to classify their purchase in one of 12 categories before clicking “enter”. (See Figure on the right – a screenshot from the app that gives consumers a choice to classify their spending).

However, this classification was introduced only recently, and the majority of consumers do not take the extra step of classifying their purchases.

For these reasons, we were unable to answer the research question on how the categories of consumption changed during the COVID-19 period in Mongolia. Nevertheless, in the table below, we provide a commentary on the likely extent of coverage of the 12 categories of spending under the VAT E-receipts system, which allows understanding differences between consumption and spending as measured by household surveys and the VAT system, respectively.

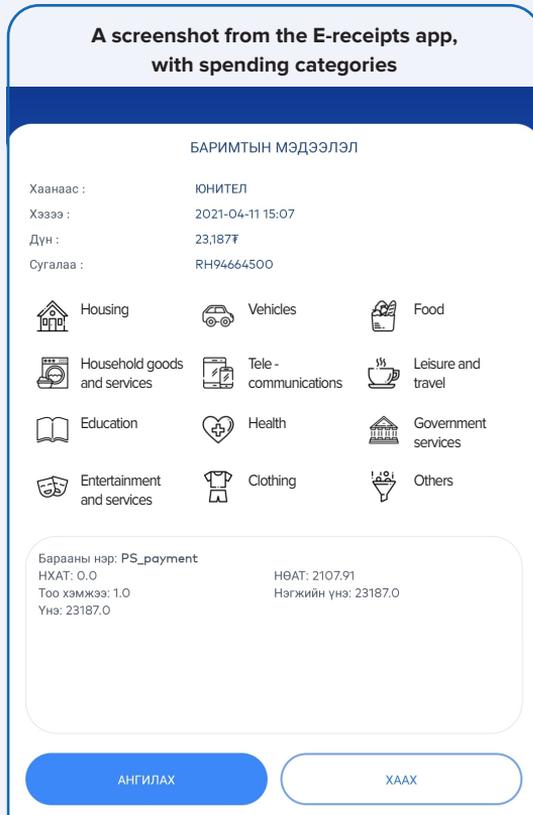


Table. The categories of spending and their coverage in the E-receipts system

Spending categories	Coverage in VAT E-receipts system
Food	<p>Generally, food items produced domestically are exempt from VAT. Therefore, consumers do not have incentives to register domestically produced food purchases in the E-receipts system, unless they are making purchases together with others – imported food and household items, for instance. In practice, the majority of purchases made via supermarkets and large multi-purpose grocery stores are registered, while food sold via informal stalls in food markets is mostly not registered.</p> <p>Self-produced and consumed food is not registered, since there are no market transactions associated with it.</p>
Clothing	<p>Clothing is taxable by VAT and thus most clothing purchases are registered in the E-receipts system. However, clothing sold by small vendors or self-produced clothing is not often registered.</p>
Household goods and services	<p>Services such as utilities, hairdressers, beauty salons, repair shops, laundry, and dry cleaning are taxable by VAT. Therefore, most such services are registered in the E-receipts system, unless they are provided by very small, informal providers.</p> <p>This category also includes durable household goods, such as household appliances, which are taxable by VAT. Since they are largely imported and sold via large retailers, most of such goods are registered in the E-receipts system. However, there is a difference between spending on household goods as measured by VAT data and household surveys. Whereas in household surveys, the expenditures on these goods are distributed over the useful life of the good to estimate consumption, in our big data sample, expenditures on such goods are registered “in bulk”, at the point of purchase.</p>
Tele-communications	<p>Telecommunication services are taxable by VAT and most of the spending on these services is registered in the E-receipts system.</p>

<p>Education</p>	<p>Education services are exempt from VAT. In addition, public secondary and primary education is free, and the majority of children are enrolled in public schools and kindergartens, although private provision of education is rising. Therefore, most education services would be not registered in the VAT system, either because there are no market transactions associated with them, or because for private education services or paid tertiary public education, consumers have no incentive to register them in the E-receipts system. Nevertheless, education service providers are required to install POS machines and enroll in the E-receipts system, and there can be a fraction of consumers who register their education payments regardless of incentives.</p>
<p>Health</p>	<p>Similar to education, health services are exempt from VAT. This includes veterinary services. However, the share of private provision of health services or paid public provision of health services is higher than in education. Health service providers enroll in the E-receipts system, and a sizeable share of consumers are likely to be registering their health-related expenditures.</p>
<p>Government services</p>	<p>Government services are exempt from VAT. Still, a minor share of consumers using paid government services are likely to be registering their expenditures on these services.</p>
<p>Leisure and travel</p>	<p>Spending on hotels, restaurants and catering services is taxable under VAT. In practice, small and informal providers of these services are unlikely to be equipped with POS machines and to be providing scannable receipts to consumers.</p> <p>Transportation services are taxable under VAT except for public transportation in the capital city. However, informal provision of transportation services is quite large in Mongolia, and therefore a significant share of spending on transportation is likely not registered in the VAT system.</p> <p>Fuels are taxable under VAT. The majority of transactions to buy vehicle fuels, such as diesel and gasoline, are likely to be registered under the VAT E-receipts system.</p>
<p>Entertainment and services</p>	<p>These services are taxable under VAT. Services provided by small and informal providers of entertainment services are unlikely to be registered in the VAT system.</p>

Housing	<p>Housing rentals are not taxable under VAT, and so are financial services.</p> <p>In addition, consumption on own housing is not imputed in our study, whereas in the HSE surveys, it is estimated by imputation.</p>
Vehicles	<p>In general, vehicles are taxable under VAT. However, purchases of used cars by businesses are not tax-deductible, so there is no incentive to register purchases of used cars.</p>
Others	<p>Most other goods and services traded domestically are taxable under VAT, with the exception of services such as financial services, funerals, cultural heritage restoration services, religious services and tourism services, and goods such as goods imported through charity or development assistance, goods produced for scientific research, exploration and demonstration, imported wood products, and imported equipment for some industries.</p> <p>In general, services and goods in the “other” category are likely to be registered in the E-receipts system if they are sold by medium-sized and large, formal sector enterprises.</p>

6.4.2. Estimation of poverty

Whereas in household surveys, poverty and inequality are estimated on the basis of consumption (although expenditure-based estimates may also be reported), in our study, they are estimated using expenditure. In other words, we use expenditure as a proxy for consumption.

Several adjustments were done to the expenditure data from the VAT system in order to make our poverty estimation based on expenditure consistent with standard statistical methodologies.

First, individual expenditure was transformed into per-person household expenditure. The big data sample was used to aggregate expenditure by individuals into households since it is household-level, rather than individual-level, consumption, that is used in estimating poverty and inequality. Then, household-level expenditures were divided by the household sizes to get per-person expenditure.⁴⁷ In other words, when a father or a mother makes purchases, they do so on behalf of the

⁴⁷ The expenditure of 23,600 individuals was aggregated into 9,826 households. Then, from the linkage between the HSES-2018 data set and the big data sample, the total number of household members was found (not only adult household members who are registered in the E-receipts system), which was 37,382.

whole household, but then, the total purchases made by the household need to be divided by the number of all family members.

Second, expenditure was adjusted for inflation. The expenditure data from the VAT big data sample for 2018, 2019, and 2020 were in current prices. We used monthly consumer price indexes (CPI) with the base of January 2018 to adjust monthly expenditure changes to measure real changes. Prices significantly differ by provinces; for instance, in the western provinces that are located farthest from Ulaanbaatar, prices tend to be higher compared with the central provinces due to high transportation costs. Therefore, we used both provincial and capital city's CPI to accurately assess living standards and expenditure values. The resulting values provide temporal and spatial inflation-adjusted per-person monthly consumption.

Once the inflation-adjusted, household-level per-person expenditure was estimated, estimation of poverty requires a poverty line. A poverty line is the sum of a food poverty line and a non-food poverty line. We used the poverty line of the HSES-2018 survey, applying it to the calculation of poverty on consumption derived from VAT data. The NSO estimated the poverty line at 166,580 MNT per capita per month at 2018 prices.⁴⁸ Both 2019 and 2020 poverty lines were derived from the HSES-2018 line, by adjusting it using food and non-food consumer price indexes for 2019 and 2020, respectively. The resulting poverty line for 2019 was 178,471 MNT and for 2020 it was 185,753 MNT.⁴⁹ Then, individuals and households with expenditures documented on their E-receipts less than the poverty line are identified as poor, while the rest are identified as non-poor in this estimation.

The calculations of poverty and inequality are then based on using individual weights to enhance the representativeness of the findings.

Since the same poverty line used in HSE surveys was applied to expenditures from the big data sample, which differs from HSE survey-based consumption in important respects, as described in the previous section, the estimates of poverty using big data will necessarily be different from those using the HSE surveys. For this reason, a more meaningful result from our analysis is the change in consumption and poverty, rather than the level.

The poverty headcount rate is given by the number of persons with income (expenditure) below the poverty line divided by the total number of persons:

$$p^0 = \frac{\sum_{i=1}^n 1(y_i < p)}{n}$$

y_i is monthly expenditure of the i -th person, calculated by dividing the expenditure of this person's household by the household size

⁴⁸ Uochi, I. (2020): Mongolia Poverty Update 2018 (English). Washington D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/532121589213323583/Mongolia-Poverty-Update-2018>.

⁴⁹ The estimated poverty lines are predicted values

p is the poverty line

n is total population and

P^0 is the poverty headcount rate.

We estimated the poverty headcount rates based on our big data set using the poverty line of the HSES-2018; for 2019 and 2020, the poverty line was adjusted (increased) for inflation.

The poverty headcount rate was weighted to reflect different probabilities of various types of households being selected in the big data set.

6.4.3. Estimation of inequality

The same per person expenditure and adjusted for inflation that was used for poverty, was also used for estimation of inequality.

We estimated the following indexes of inequality.

The first one is the Gini coefficient, given by:

$$G = \frac{1}{n-1} \left(n+1 - 2 \times \frac{\sum_{i=1}^n (n+1-i)y_i}{\sum_{i=1}^n y_i} \right)$$

where n is population and y_i is the expenditure per person of the i -th person.

The second measure of inequality is the Theil index.

$$T_T = T_{\alpha=1} = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\mu} \ln\left(\frac{y_i}{\mu}\right)$$

Where N is population, y_i is the expenditure of the i -th person and μ is the mean expenditure. The natural logarithm transformation makes the change in inequality more sensitive to the increase in the share of the richer persons' share in total (national) expenditure beyond a certain threshold - beyond this threshold, inequality rises more steeply than the share of the richer persons' expenditure in total.

The third measure of inequality is the Palma ratio – the ratio of expenditures of persons in the richest decile in the big data sample to the expenditure of the poorest 40 percent, given by:

$$Palma_index = \frac{\sum_{i=1}^n y_i (y_i \in percentile_{90}^{100})}{\sum_{i=1}^n y_i (y_i \in percentile_1^{40})}$$