***Construction of Admin Sources-Based Household Frame Using GIS for Abu Dhabi Emirate***

Ali. R, Dr. Al Refaei. M, Al Dhaheri. A, Al Dhanhani. A, Al Suwaidi. M, Al Mazrouei. F

Statistics Centre - Abu Dhabi

**Abstract**
Census data on families and households contains counts of households, data on the distribution of household sizes, and details on the relationships between people within households. Government decisions such as income support and social housing are developed and evaluated using this data. Currently, the Statistics Centre in Abu Dhabi (SCAD) produces household and family information by conducting household surveys through field applications. Many government agencies place a high value on information regarding houses and families. Administrative data gathered for purposes other than producing official statistics, provide new ways of collecting data on individuals, households, and dwellings. This paper aims to construct an admin sources-based household frame for the Abu Dhabi Emirate. In this study, property rental contracts data and water and electricity bills data were used. Further, both datasets were filtered out for Jan 2022, and only the residential category was selected. **Though both datasets represent the households, to enrich the spatial and attributes coverage both datasets were spatially joined, and the spatial duplicate records were excluded. The researchers have calculated the difference between the admin-based data and survey data to assess the quality. As a result, we have found that there is a difference of 5% - 10% in 17** **enumeration areas, a difference of 10% - 20% in 26 enumeration areas, a difference of 20% - 50% in 62 enumeration areas, and 77 enumeration areas suffer from a huge difference of more than 50%.**

**Introduction**

Researchers and policymakers can use household frame data to analyze trends, conduct surveys, and develop statistical models to better understand the population of a specific geographic area. It can help businesses develop targeted marketing campaigns and reach their intended audience more effectively. A household frame is a critical tool for emergency response efforts, such as in the case of natural disasters or public health emergencies. It can be used to identify households in need of assistance and to allocate resources to those areas more quickly.

In recent years, Geographical Information Systems (GIS) have become increasingly popular in their ability to collect, manage, and analyze spatial data. The application of GIS in developing household frames using administrative purposes has proven to be an effective and efficient approach. This research paper aims to explore the potential of using administrative data and GIS technology to develop a reliable and accurate household frame. The objective of this study is to create a list of households with at least the following attributes:

* Geographical location of household
* Name of household head
* Phone Number
* Nationality
* Emirates ID

Household frames can be derived from property rental contracts and water and electricity bills because these datasets contain information about the occupants of a household and the property they occupy. In property rental contracts, landlords typically include information about the tenants who will be living in the property. This can include the names of all tenants, their contact information, and the duration of their lease. This information can be used to create a household frame that includes all individuals living in the rental property. Similarly, water and electricity bills can provide information about the occupants of a household. When someone opens a utility account, they are typically required to provide information about the occupants of the property, including their names and contact information. This information can be used to create a household frame that includes all individuals living on the property. Overall, household frames derived from property rental contracts and water and electricity bills can provide valuable information about the composition of households and the individuals who occupy them.

Building’s footprints can be used to build a spatial link between water and electricity bills data and rent contracts data. The reason to use rent contracts and water and electricity data together is to enrich households' information by combining attributes and to have more coverage in case the household is available in rent contracts data but not available in water and electricity data.

Studies review several methods of linking administrative and survey data. Survey data are used for many purposes and are one of the most important sources of information for policymakers and researchers. The article “Assessing Survey Data Quality Making Use of Administrative Data” by B. Felderer highlights the importance of survey data in informing policy decisions and the need for such data to be accurate and reliable. It discusses that the use of administrative data can help to mitigate some of the errors and improve the overall quality of survey data. The article’s chosen methodology provides suitable results by analyzing correlates of survey nonresponse and measurement error. The samples of the surveys that were analyzed were drawn from administrative records, so they had useful information on survey respondents and non-respondents. As a result, all measures used to increase response rates can in principle affect nonresponse bias. Administrative records give a great opportunity to study this issue. However, the author suggests that a careful assessment of the available data sources and methods for integrating them with survey data is necessary to ensure that the resulting data is accurate, reliable, and informative (Felderer, 2015).

Another review has been done exploring the article “Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net” by Bruce D. and colleagues. The study examines the consequences of survey underreporting of transfer programs for prototypical analyses of low-income populations by implications of CPS survey errors, showing that they can lead to large errors in the analyses of three kinds of literature that rely on program receipt and income data. The authors replace survey responses on the receipt and amount of government transfers with administrative records for four income transfer programs (food stamps, TANF, General Assistance, and subsidized housing) over a four-year period. As a result, in most cases, the administrative records are extremely accurate (they contain actual payments made, they are validated by the agency, and their definitions are comparable to survey definitions). The administrative data are linked to survey data at the household level with a high match rate because validated Social Security numbers are required to receive three of the programs. The authors highlighted the limitations of traditional survey data in accurately measuring income and poverty. They argue that survey data are subject to reporting errors, underreporting of income, and measurement errors, which can lead to inaccurate estimates of poverty and inequality. The review then explores the potential benefits of using linked surveys and administrative data to overcome these limitations and can provide more accurate and comprehensive information on household income. The article underscores the need for careful consideration of the challenges, and limitations associated with using linked data to measure income and poverty (Bruce D. Meyer and Nikolas Mittag, 2019).

**Study Area**

The study area for this research is Abu Dhabi Emirate (Figure 1) located in the geographical area between 24° 54° 22' 0.0084'' E and 24° 28' 0.0012'' N. Abu Dhabi is located on the eastern coast of the Arabian Peninsula, bordering Saudi Arabia to the south and Oman to the east. The emirate covers an area of approximately 67,340 square kilometers, with a coastline extending over 700 kilometers along the Arabian Gulf. Abu Dhabi Emirate has three regions: Abu Dhabi, Al Ain, and Al Dhafra. According to the latest estimates from the Abu Dhabi Statistics Centre, the population of Abu Dhabi was 2,161,619 people in 2011. Of this total, the majority (60%) were living in Abu Dhabi Region (Final Population Estimate Mid-year 2011, 2013).

The Emirate of Abu Dhabi has three Municipal Regions. Abu Dhabi region encompasses the capital city of Abu Dhabi and its surrounding areas. It is the political, economic, and cultural center of the emirate. The region is characterized by a rapidly expanding urban landscape, major infrastructure projects, and diverse economic sectors, including finance, tourism, and industry. (Abu Dhabi Statistics Centre, 2021). Al Ain is a region within the Emirate of Abu Dhabi that has a fascinating history, diverse culture, and breathtaking natural scenery, and it is the garden city of the UAE. The city is renowned for its many historical and cultural sites of importance. Al Dhafra is the western region of Abu Dhabi Emirate. It covers the largest area, constituting 71% of the emirate's total area. Despite its vast size, the population density is very low, making it the region with the smallest population. A significant feature of this region is its natural resources, particularly the gas and petroleum reserves, which account for 90% of the Emirate's reserves and 90% of the country's reserves, which are vital for the local economy (Oxford Business Group, 2016).



Figure 1: Abu Dhabi Emirate Regions.

**Data Sources**

**Building’s Footprints Data:**

Building Footprints are polygon features representing the real-world building outline captured using satellite images with a description of the exact size, shape, and location of the building.

**Administrative Data:**

*1. Water and Electricity Bills*

A dataset containing observations that attribute water and electricity consumption. This consumption is reported on a monthly basis, water consumption is presented in cubic meters (m^3), while the electricity consumption is presented in kilowatts per hour (kwh).

*2. Rent Contracts*

A dataset containing all rent contacts in the Abu Dhabi Emirate. It has three tables: Contracts, Units, and Villas.

Below table shows the variable’s coverage for the administrative data. Water and electricity bills data have records of the monthly transactions and repeat twice if the households have both electricity and water connection.

Table 1: Water and electricity bills data variable’s coverage.

|  |
| --- |
| **Database’s variable completeness during Jan 2022** |
| **Data Source** | **Variable Name** | **Percentage of Completeness** |
| Water and Electricity Active Connections Jan 2022 | SA\_TYPE | 100% |
| CUSTOMER\_CLASS | 100% |
| PREMISE\_TYPE | 100% |
| PREMISE\_ID | 100% |
| EMIREATES\_ID | 63% |
| NAME | 99% |
| PHONE | 98.4% |
| NATIONALITY | 64.7% |
| NATIONALITY\_CODE | 8.7% |
| TARIFF (unit type) | 97.9% |
| Rent contractsActive ContractsJan 2022 | UNIT\_USAGE\_TYPE | 97.1% |
| PROPERTY\_USAGE\_TYPE | 100% |
| PREMISE\_ID | 20% |
| TENANT\_EMIRATE\_ID | 11% |
| TENANT\_TRADELICENSE\_NO | 11.5% |
| TENANT\_EMIRATE\_ID OR TENANT\_TRADELICENSE\_NO | 22.6% |
| TENANT\_NAME | 22.6% |
| TENANT\_PHONE | 19.8% |
| TENANT\_NATIONALITY | ATTRIBUTE NOT AVAILABLE |

**Filed Data (Survey):**

1. *Enumeration Areas*

Enumeration areas (EAs) are geographic subdivisions or units defined by statistical agencies for the purpose of conducting data collection activities. These areas are designed to facilitate efficient and systematic enumeration of households and population groups. Enumeration areas are typically defined based on geographic boundaries, such as neighborhoods, blocks, or specific geographic features. They are used to divide a larger geographic area into smaller, manageable units to ensure comprehensive coverage and accurate data collection. Enumeration areas help in organizing field operations, assigning enumerators, and ensuring that all households within a given area are covered during data collection exercises like censuses or surveys.

1. *Household Data*

Specific information collected from individual households. This data encompasses a wide range of variables related to the composition, characteristics, and behaviors of households. It includes demographic information (such as household size, age, and gender distribution), socioeconomic indicators (such as income, education, employment status), housing conditions, and other relevant data points. Household data is collected through various methods such as surveys, censuses or interviews.

**Methodology**

The methodology obtained to construct the admin-base household frame is highlighted in Figure 2 below and explained in the following steps:

1. Both water and electricity data & contracts data were filtered out for Jan 2022 and filtered out for the residential category
2. Datasets were displayed on the map using locational fields
3. Building information attached to both datasets using spatial join, to get building info as linkage key to join both datasets together
4. In case the same household appeared in both rent contracts and water and electricity data, then both records were joined to reduce duplication and enrich information for that household
5. In case the household record is available in water and electricity data but is not available in rent contracts data, then it is included in the result, to give more spatial coverage on the household regardless of missing respective attributes
6. In case the household record is available in rent contracts data but is not available in water and electricity data, then it is included in the result, to give more spatial coverage on the household regardless of missing respective attributes
7. Produced household counts for each enumeration area
8. Household count of each enumeration area is compared with the household count from the household field survey 2021 to assess the quality of the admin-base household frame

******

Figure 2: Methodology Diagram.

**Results**

After testing and comparing the data of the household frame update 2021 (HHFU) collected from the field, with the administrative data (water and electricity bills and rent contracts), the comparison was made (Figure 3) at the level of 198 enumerated areas in Abu Dhabi and showed the following:

* The total number of households in all enumeration areas in the HHFU data in 2021 is 29,897 households, while the administrative data in those areas is 22,945 households.
* At the level of enumeration regions, some regions had more administrative households than the number of households from the field, and other regions were on the contrary. The following are the final observations found:
	+ In 16 out of 198 enumeration areas, the range of differences between the number of field households and the administrative households did not exceed 5%
	+ 17 enumeration areas where the range of differences between the number of field households and the administrative households ranged from plus or minus (5% - 10%).
	+ 26 enumeration areas with a range of plus or minus differences (10% - 20%).
	+ 62 enumeration areas with plus or minus difference range (20% - 50%).
	+ 77 enumeration areas where the differences between the number of field households and the number of the administrative households exceeded plus or minus 50%.

Figure 3: The comparison between admin data (Rent Contracts, Water and Electricity bills) and household frame update 2021 (HHFU).

When making a comparison, there are aspects that must be considered, namely:

* The household frame update is a field survey that took place within a specified reference period in (the third quarter of 2021), while the administrative data reference period is current.
* Also, the process of data collection in the field is naturally permeated by some errors, lack of coverage and response, and other errors.
* Within the water and electricity database, there are 954 electricity subscribers, each of whom has more than one electricity subscription, where there are subscribers, each of whom has 10 accounts, others whom each have 12 accounts, and so on...
* The rent contracts data file includes around 40 units that describe their use (grocery store, exhibition, shopping center, etc.) even after applying the residential filter in the first place. There must be a way to limit these cases and exclude them from the database.

There are clearly significant differences between the number of field households and the number of administrative households from the three sources at the level of enumeration areas. Through the above data, there is no indication of which of the sources is more accurate. The data of the household frame, as a field survey, includes percentages of errors in accuracy, coverage, collection, and time reference, and the data of the administrative data also includes aspects of bias.

**Conclusion**

Using field data to validate administrative data is an important step towards improving the quality of the produced official statistics and the accuracy of policy decisions. While administrative data sources have many benefits, they are subject to errors and inconsistencies that can lead to incorrect conclusions. Field data collection provides a means of verifying and validating administrative data. However, based on the foregoing of our study, the administrative datasets need to be verified for their efficiency and completeness in terms of:

* All subscriptions of water and electricity are actually used for residential purposes, and the actual beneficiaries are the tenants.
* The recurring subscriptions of one individual are actually used by that individual, and not by other individuals or families.

In conclusion, the field survey remains an essential means of verifying and validating administrative data, but it is also necessary to consider the limitations and potential errors that may arise during data collection. Ultimately, the use of both administrative data and field surveys is crucial for producing accurate official statistics and informing effective policy decisions. This paper served as a preliminary pilot study aimed at gaining insights into the potential benefits of utilizing administrative data. Within our organization, we are considering a shift towards adopting administrative-based registers instead of relying solely on field surveys. Through this pilot, we have discovered significant opportunities for enhancing our methodology, which could lead to improved results and outcomes in the future.

***Bibliography***

Abu Dhabi Statistics Centre. (2021). *Statistical Yearbook of Abu Dhabi*. Retrieved from https://www.scad.gov.ae/en/Pages/default.aspx.

Bruce D. Meyer and Nikolas Mittag. (2019). Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net. *American Economic Journal*, 176–204.

Felderer, B. I. (2015, September). Assessing Survey Data Quality Making Use of Administrative Data.

Statistics Centre Abu Dhabi (2013). *Final Population Estimate 2011.*

Oxford Business Group (2016). The Report Abu Dhabi 2016.