Paper 5163-2020

An Astounding Lack of Data

Christopher Hooks, North Carolina State University

ABSTRACT

In today's world, where there are quintillions of bytes of information generated every day, we still suffer from a lack of data. Data is driving the decision-making in business, but what about governments? Why are cities behind when it comes to data-driven decision-making? This paper looks at the availability of city-level data and how it can be used in decision-making at the local level. The topic is further extended with a discussion of existing barriers to using data for informed decision-making in local governments.

INTRODUCTION

The goal of this paper is not to introduce any data, but rather to introduce a theoretical model with the goal of creating sustainable cities for the future. The developments within information technology over the past few decades have forever changed the landscapes of the local economy. It has never been easier for a company to manage multiple physical sites across the globe. This is because a company is internally cooperative in order to compete more effectively externally. Cooperation within companies has been improved by the developments in information technology. Workers on the other hand face increased competition, not just locally, but globally as well. Companies can easily find substitute workers for almost any job they offer. This has given large national and multi-national corporations a competitive advantage, increasing their ability to acquire market share through outcompeting or, more commonly, buying up competition.

The Federal government has put a major emphasis on ensuring continued economic growth during the past century. In the 1930s, Keynes' theories on economics began to be adopted by many governments, and the United States was no exception. Under Keynes' theories, fiscal policy can be used to manage the business cycle fluctuations of the economy. Deficit spending during a recession can boost demand and thereby employment. In the 1980s, President Reagan cut taxes and adjusted budget allocations. In his 1981 address to a joint session of congress, Reagan signaled a shift to Supply side economics when he stated, "The taxing power of government must be used to provide revenues for legitimate government purposes. It must not be used to regulate the economy or bring about social change." More recently, President Trump's tweet that signaled a positive jobs report more than an hour before the official release was even called into some question because of its potential to influence financial markets. The government's attitudes and interactions with markets do have an impact. The government's policies do impact the economy and socioeconomic status of its citizenry. (This is not meant to be a full history.)

The advancements in information technology have had a disruptive effect on the economy. The ability to coordinate activities across the globe has given large corporations a competitive advantage. However, this advantage has also coincided with a rise of income inequality. The market power of these firms has risen while there has been a decline in the market power of smaller firms and consumers. Consider the labor market situation in the

19th century. The terrible conditions of the working class led to the rise of Labor Unions. These unions brought about positive changes because they increased the market power of the individual worker through collective bargaining. In much the same way, cities can cooperate to increase their overall market power through increased coordination of activities. Cities can implement change through utilizing the tools we already have available. By so doing, they can greatly influence their future courses. This would take strong initiative on the part of cities because there is not an out-of-the-box solution that will work.

We must shift from national-centric to community-centric ideology in economics. This means that data collection and its processes must be understood and managed by the local stakeholders. Comparisons between countries and statistics given at the national level ignore the heterogeneity that exists between geographic locations. Communities create more comparable groups across the globe, allowing the data to provide greater levels of intelligence. This intelligence would enable municipalities to individually manage their own economies more efficiently and interact with each other in more environmentally sustainable ways.

How well do cities currently understand the importance of data or the development of digital services in interactions between the local government and its citizenry? While there are definitely cities in the United States that have mobile apps that enable citizens to access specific information regarding traffic/transit, news, politics or healthcare services, the interaction between city government and its citizenry involving digital (especially mobile) technologies is still in its infancy. Most cities do not even have any data with API access, and their digital technology skills are desperately lacking. The City Open Data Census currently ranks Cincinnati, OH and Las Vegas, NV as the most open data cities in the United States. The survey currently has 226 places(cities) and 880 datasets listed, 471 of which have open access. These are not promising numbers given that within the 50 states (excluding territories) there are 3, 142 counties alone.

A COPERNICAN REVOLUTION IN ECONOMIC ANALYSIS

Since the 1930s, economic thinking has centered around macroeconomic policy and decision making. Our focus has been on real business cycle fluctuations and the use of monetary and fiscal policy to "minimize the variability" of GDP growth and unemployment. Our focal point has been the national economy, and our data collection has matched these economic theories.

I propose a simple change to this approach. Data collection, warehousing and processing should happen at the local area (community level) first. There is an astounding lack of accessible, localized, time series data sets for researchers and economists. There are many privacy concerns with collecting more detailed economic data at the local level, but these concerns must be addressed in order to promote sustainable economic growth and a more equitable distribution of resources.

Currently, you can get annualized economic data for counties, micropolitan areas and metropolitan areas from 2001 through 2018 (as of February 2020) from the Bureau of Economic Analysis. For the Census Bureau's County Business Patterns, data are available on an annualized basis from 1946 through 2017. This data has seen a few changes over the years, and data from early years has a few inconsistencies. Beginning in 2017, sectors with fewer than 3 establishments are omitted, but this is the most consistent source of micro level economic data. County level employment data by sector and month, and total wages

by sector and quarter are available from the Bureau of Labor Statistics beginning in 2015 to present with a 2-3 quarter lag on availability. In contrast, many data series are available on a quarterly and monthly basis for the national economy. However, for local-level economic policy, this national data is not effective, as all geographic units are not homogeneous.

Our macro theories have been based on national level data that yield poor results. We have compensated for this lack of data by doing country level comparisons based on per capita data, but this is still like using a flip phone camera on a satellite. While you will capture all of the data, you have no capacity to zoom in. Can you imagine Instagram with flip phone cameras? Comparing geographic regions on a per capita basis is more of an apples to oranges than an apples to apples comparison.

So how do we make an apples to apples comparison more feasible? If our cities and towns had their own data management and analytics departments, and this data was readily available to researchers and data scientists, what would our society look like? Would cities be able to effectively utilize this data to make better policy decisions? Businesses understand the importance of detailed data when it comes to decision making, so why the stark contrast when it comes to cities?

Our focus in the past has been on Federal programs to collect and aggregate data at various levels of geographic detail. With the innovations in digital technologies over the past several decades, our ability to collect more granular data has increased, but our government data collection processes have lacked luster by comparison. That is not to say that our government does not have the capabilities to collect high quality data with great definition, we have just focused most of our data collection efforts on national security while economic data has fallen short by comparison.

The knowledge-based economy is continuing to grow in importance across the globe. It is essential to our continued economic vitality to focus more effort on collecting geographically specific economic data in much greater detail. Doing this will improve decision-making capabilities at all levels of government. From an economic perspective, data is the new gold. From a societal perspective, the ownership, accessibility and utilization of this data will have enormous consequences in the coming decades. These consequences can be good and/or bad, but the more asymmetric the data, the more inequality will exist within our society, and the more eroded our democracies will be.

A brief review of global history will show that the concentration of power and inequality are positively related. It will also show that these societies tend to become unstable and end in civil wars. Therefore, it is extremely vital to our future generations that we give this topic careful consideration and plot our course ahead with care. Our decisions on data rights and privacy will have real world consequences.

Public opinion towards data ownership and privacy rights are still in a rather embryonic state. Many people feel that they do not have anything to hide and so they do not care as much about protecting their data, but at the same time, they do not realize the extent of the information that is collected about them on a daily basis because most of it is collected without their explicit knowledge. A 2018 article in the MIT Review put it this way:

An incremental erosion of privacy is tough to notice and does little harm to anyone—just as trace amounts of carbon dioxide are scarcely detectable and do no environmental harm to speak of. But in the aggregate, just as large amounts of greenhouse gases cause fundamental damage to the environment, a massive shift in the nature of privacy causes fundamental damage to the social fabric. (Tisne, 2018)

The importance of data in decision-making cannot be overstated. However, the way that this data is accessed and the impacts of this data on society must be carefully monitored. An individual has little ability to actively protect their data from exploitation. Therefore, there must be systems in place to protect these individual rights. This is the most important factor to consider in establishing new data collection protocols for generating localized data sets. While this is a highly complex topic, it should not be avoided because of its complexity, nor should it cause citizens and municipalities to shy away from establishing new data collection systems. If municipalities are to be effective in the 21st century, it is absolutely essential that they adapt their laws and policies to reflect the ongoing changes in technology.

AN INVESTMENT PORTFOLIO APPROACH TO THE LOCAL ECONOMY

We all know the importance of having an investment portfolio for retirement. We wish that we could take advantage of insider trading knowledge to win big, and then we could just transfer our new-found wealth into a riskless asset to wait for the next insider opportunity. Unfortunately, there are no crystal balls that let you legally do this. This is why we have portfolios of investments. Nobody likes risk, but everyone loves high returns.

For a municipality, this should be no different. A portfolio allows us to optimize our returns based on our individual risk tolerance. This is simple basic knowledge. Why then are we not looking at the real physical investments within our cities using a similar approach? We all know the story of Detroit and the decimation of the local economy with the influx of foreign cars. How do we engineer our cities to mitigate the risk of such an occurrence, while also considering the locally specific factors that create comparative advantages in specific sectors?

In portfolio theory, investors seek to maximize their return to risk ratio. Returns can be increased by heavily weighting our investments in specific sectors, just as Detroit and the automobile industry. However, this type of return-maximization-only strategy leads to higher variance of returns for a municipality as a whole. The economy in Detroit was primarily comprised of layers of production in a single industry, and when that industry suffered from external competition, the entire area suffered a massive, negative shock, which it is still working to recover from. In order to limit the risk from large negative shocks, investors distribute their investments across multiple firms and sectors. A municipality can likewise limit their risk by promoting a diversified economy. These two forces are counteracting. As diversification reduces the variability of returns, it also reduces the potential for high returns, both positive and negative.

In order for a municipality to actively manage its economic portfolio, it must be able to access the data generated by the individual economic actors (both firms and individuals) in the local economy. This data would then be used to target specific areas of the economy for development. Just as investors buy and sell investments to alter their exposure to different

sectors. In a similar sense, municipalities can selectively target economic development funds to adjust their economic portfolios.

DETERMINING THE LEVEL OF ECONOMIC ACTIVITY

The total level of economic activity in a municipality can be viewed as a sum of the value of all transactions made during the specified time period within the defined geographic area. Each transaction represents a trade that occurred where the supplier provides the customer with a good or service. In exchange, the customer trades another good or service (usually cash) with the supplier. This transaction is recorded by the supplier, possibly the customer, and any intermediary facilitators, such as a financial institution when the customer uses a debit or credit card.

The municipality is faced with the issue of determining the level of economic activity for each business within its jurisdiction. The municipality can use sales tax records to extrapolate the level of transactions for businesses that collect sales tax, but not all businesses collect sales tax. This data itself would only give information on the volume of cash flows and may not be of greatest benefit to economic planning efforts. Municipalities can also access aggregated data from the IRS by zip code, but the most recent data as of February 2020 is from 2017. The IRS data does not distinguish between sectors for business returns either. Essentially, there is no effective source or system for a municipality to gather and analyze economic data that is timely enough to make effective economic policy decisions.

Imagine if you constantly made your day-to-day decisions based on what happened 2 years ago. Your decision-making capabilities would be severely limited. In contrast, the national-level quarterly GDP estimates, Monthly Personal Income and Outlays, and Monthly Jobs reports are available within days to a few weeks after the end of the period. These reports provide information that drives decision making not only in the United States, but around the globe. If this type of information were made available at the community level with the same frequency, municipalities could drastically improve their decision-making capabilities.

DEFINING THE SYSTEM

Since timely, accurate data is not readily available for municipalities and is not collected in any systematic way, it offers the opportunity to design an entirely new system to accomplish this. In order to create an optimal system, it is important to consider the underlying data generating processes. In the modern exchange economy, almost all transactions are associated with an identical cash flow in the opposite direction as the flow of the good or service. Therefore, the most granular data would be collected at this level. For the sake of argument, consider this like you would a physics experiment in a lab setting and assume friction is negligible.

As an example within a retail setting, imagine a customer who purchases a random basket of goods. The company will normally provide the customer with a printed receipt that itemizes the goods or services purchased, price per unit, the total, and the added taxes. If the company has an inventory tracking system, it will automatically update the inventory records for the items just purchased. If the customer pays with cash, there is no other record of this transaction recorded. If the customer uses either a debit or credit card, there is a transaction record generated that records the merchant details and the amount of the transaction, and the customer's account balance is updated. However, there is no information regarding the actual products or services purchased.

If each transaction that occurred generated a record that included: the origination of the goods and services, type of good or service, quantity, price, total value of the transaction, and customer information including the bank, location of residence, and basic demographic data, then a nearly perfect representation of the cash flows within the geographical area of a municipality could be generated. In reality, there is no new data being recorded, but the quality of the information would be dramatically improved. The real challenge would be accomplishing this type of data collection in an ethical manner.

In order to convince members of a municipality to cooperate in a system that collected and recorded data on every transaction, there would need to be substantial benefits. These benefits would also need to be weighed against the costs and inherent risks of the system. The ownership of the data, openness of the data, uses of the data, restricted accessibility of the data, and the security of the data are areas of primary concern for consideration.

While the primary goal of the municipality would be to maximize the risk/return ratio of the area's economic portfolio, the choice of metrics used to measure the return and risk would likely lead to some heated debates. Some of the main factors that the municipality could focus on would include output, employment, income, inequality, pollution, and even arbitrary measures of "the quality of life". The choice of these metrics could be chosen by the individual municipality via any selection process. The consequences of the choice of metric would theoretically lead to different socioeconomic paths, and municipalities with the best optimization strategies would eventually dominate municipalities with less optimal strategies. However, domination should not be assumed to be business related. There are many other factors in our society besides GDP. An optimal strategy should seek to maximize the benefit of the society as a whole without creating losses in other societies. In economic speak, this would be the maximization of the social welfare of the municipal region within a global setting.

In order to facilitate broader aggregation of data at the regional, state and national levels, various aggregated data could be made available via a data sharing cooperative network. This network would take the place of the current Federally administered data collecting agencies. The costs of the data collection, warehousing and processing would be distributed across the individual municipalities. In addition, an independent agency would need to be established that would be responsible for ensuring that the data collection processes followed standardized protocols, verifying the authenticity and quality of the data, and enforcing standards. This agency could be funded by a population-based fee with additional revenue generated from fines. To ensure the long-term viability, an endowment for funding should be instituted to compensate for the variability of annual revenues and expenses. Control of this agency should be divided between both appointed and elected individuals. A fully designed oversight agency is beyond the scope of this paper, but it should be designed in such a way that it limits the possibility of exploitation through corruption.

For firms within a specific municipality, the benefits of the system would disproportionately benefit locally owned small and medium sized enterprises (SMEs) because it would give them access to market data, financial analysis, inventory management and other ancillary business tools that they would normally not be able to afford. Large firms normally have the financial resources to self-fund these endeavors and make it more difficult for SMEs to operate in a highly competitive market due to asymmetric information. By improving the symmetry of information in the market, large firms will face increased competition leading

to more socially optimal market outcomes (i.e. more firms, lower prices, and greater quantity of supply).

Individual firms within a municipality would be able to see the data generated by their own business, but they would also be able to see locally aggregated market information. Information about specific product's demand predictions, automated product order generation and inventory control, cash flow, capital budgeting analysis and simulation tools would be standard services provided as a benefit of participation. These services could be tailored to suit the individual needs and preferences of the firm.

Individuals within the municipality would benefit from having more detailed records of their purchases. This would facilitate the personalized budgeting tools including financial situation simulators that could be used to simulate the impact on their individual cash flows when considering purchases. Product category spending limits and/or alerts could also be established in a similar way as one can set screen time limits for the different apps on their smartphones. Additionally, consumers would benefit from reduced search times. This is the main attraction of online retail giants such as Amazon. In the 1990s, there were websites for individual companies and individual product categories, but the time required for customers to search through all of these websites before making a purchasing decision was similar to driving to local stores, especially when considering the additional wait time required for shipping. Amazon successfully exploited these inefficiencies by hosting multiple products on the same site and offering rapid delivery services.

For the municipality as a whole, there would be a substantial increase of investments in information technology infrastructure. This would also lead to the growth of jobs within this sector of the economy. Jobs related to computer science and data analytics would also be a significant source of new jobs and income for many municipalities. Ownership of the information technology infrastructure should be limited to within the municipality as this sector is an underlying driver of all other sectors in the economy today. This would guarantee that the local populations would have control over the quality of infrastructure that they have access to. Currently, this infrastructure is primarily controlled by a small handful of companies who have little profit incentive to improve their infrastructure as a result of antiquated, geographic monopoly agreements and regulations. The financial flows from profits generated by this system would be fed back into the local economy instead of being extracted from the local municipality, as with the current model.

Secondly, the cooperative ownership and development of the system within the municipality would drive further growth of locally owned businesses due to their improved ability to compete with large national and multinational firms. The municipality can improve this growth through individually targeting specific sectors and subsectors within the economy, maximizing the recirculation of cash within the local area. Nearby municipalities also have the opportunity to cooperate and organize production in a way that minimizes total environmental costs. Cooperation and coordination in the production of goods and services at a regional level can reduce transportation costs, improve the feasibility of recycling, support the use of more recyclable materials, and thereby improve the ecological sustainability of the region.

CONCLUSION

The need to access local economic data is vital for the future management of the "national" economies. This can occur through a myriad of paths, but most definitely, it will eventually occur. It may be the national government, business, or any lower level of government that starts the process, but who will win? If the collection of this data is left to the federal government, then the likelihood of municipalities having timely access is still insignificant. If corporations take the responsibility for collecting this data, then the municipalities will likely have a tiered access depending on their willingness to pay. If the municipality begins the process, they will have the maximum opportunity to benefit from it.

Some of the main questions society must answer though are: What is the role of government going to be in the 21st century? Will large national governments rule or will there be a resurgence of power in the local governments? Should government actively seek to improve societal and economic outcomes? What is the proper role of government? What is the role of large national and multinational corporations? How will we navigate these murky waters of Data Rights and Data Privacy? How will we transition into a more ecologically sustainable economy, or will we simply and ignore the looming environmental concerns while we exhaust our limited resources? These may be broad questions, but humanity must address these concerns. A primary way of addressing these is by shifting our emphasis on the national to the local economy and focusing on making them self-sustaining and environmentally friendly.

REFERENCES

Address Before a Joint Session of the Congress on the Program for Economic Recovery. (1981, February 18). Retrieved from https://www.presidency.ucsb.edu/documents/address-before-joint-session-the-congress-the-program-for-economic-recovery-0

Knowledge, O. (n.d.). - U.S. City Open Data Census. Retrieved from http://us-cities.survey.okfn.org/

Pinzón, D. F. D. B., & Souza, F. T. D. (2016). A data based model as a metropolitan management tool: The Bogotá-Sabana region case study in Colombia. *Land Use Policy*, *54*, 253–263. doi: 10.1016/j.landusepol.2016.02.019

Ribeiro, J., Silva, P., Duarte, R., Davids, K., & Garganta, J. (2017). Team Sports Performance Analysed Through the Lens of Social Network Theory: Implications for Research and Practice. *Sports Medicine*, *47*(9), 1689–1696. doi: 10.1007/s40279-017-0695-1

Souza, J., Francisco, A., Piekarski, C., & Prado, G. (2019). Data Mining and Machine Learning to Promote Smart Cities: A Systematic Review from 2000 to 2018. *Sustainability*, *11*(4), 1077. doi: 10.3390/su11041077

Tisne, M. (2018, December 18). It's time for a Bill of Data Rights. Retrieved from https://www.technologyreview.com/s/612588/its-time-for-a-bill-of-data-rights/

Visvizi, A., Lytras, M. D., Damiani, E., & Mathkour, H. (2018). Policy making for smart cities: innovation and social inclusive economic growth for sustainability. *Journal of Science and Technology Policy Management*, *9*(2), 126–133. doi: 10.1108/jstpm-07-2018-079

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Christopher Hooks
chooks2@ncsu.edu

BASIC INSTRUCTIONS

WRITING GUIDELINES

Trademarks and product names

To find correct SAS product names (including use of trademark symbols), if you are a SAS employee, see the Master Name List. Otherwise, see SAS Trademarks.

- Use superscripted trademark symbols in the first use in title, first use in abstract, and in graphics, charts, figures, and slides.

- Do not abbreviate product names. For example, you cannot use "EM" for SAS® Enterprise Miner™. After having introduced a SAS product name, you can occasionally omit "SAS" for certain products, provided that your editor agrees. For example, after you have introduced SAS® Simulation Studio, you can occasionally use "Simulation Studio."

Writing style

- Use active voice. (Use passive voice only if the recipient of the action needs to be emphasized.) For example:

  The product creates reports. (active)
  Reports are created by the product. (passive)

- Use second person and present tense as much as possible. For example:

  You get accurate results from this product. (second person, present tense)
  The user will get accurate results from this product. (future tense)

- Run spellcheck, and fix errors in grammar and punctuation.

Citing references

All published work that is cited in your paper must be listed in the REFERENCES section.

If you include text or visuals that were written or developed by someone other than yourself, you must use the following guidelines to cite the sources:

- If you use material that is copyrighted, you must mention that you have permission from the copyright holder or the publisher, who might also require you to include a copyright notice. For example: "Reprinted with permission of SAS Institute Inc. from *SAS® Risk Dimensions®: Examples and Exercises*. Copyright 2004. SAS Institute Inc."

- If you use information from a previously printed source from which you haven't requested copyright permission, you must cite the source in parentheses after the paraphrased text. For example: "The minimum variance defines the distance between cluster (Ward 1984, p. 23)

TIPS FOR USING WORD

To select a paragraph style

1. Click the HOME tab. The most common styles in your document are displayed in the top right area of the Microsoft ribbon. If you don't see a style that you want, click the slanted down arrow at the bottom right corner of the Styles area, and scroll through the list. The main styles for this template are headings 1 through 4, PaperBody, and Caption. Avoid using other styles.

2. To change a paragraph style, click the paragraph to which you want to apply a style, and then click the style that you want in the ribbon.

3. PaperBody (used for most text) is automatically applied when you press Enter at the end of any heading style or the Caption style.

To insert a caption

1. Click **REFERENCES** on the main Word menu.

2. Click **Insert Caption**.

3. Select the **Label** type that you want.

4. Click **OK**.

To insert a graphic from a file

1. Click **INSERT** on the main Word menu.

2. Click **Picture**.

3. In the Insert Picture dialog box, navigate to the file that you want to insert.

4. When the name of the file that you want to insert is displayed in the **File name** box, click **Insert**.