APPENDIX A

INDUSTRIAL CITIES INITIATIVE

Edited by Susan Longworth

Appendix A: Overview of key data sources and compilation methods

[1] U.S. Census Bureau

The U.S. Census collects information on the American population and housing every ten years for use in policy-making and research. Until recently, it was distributed in two forms: a short form that counts all residents as mandated by the Constitution, and a long form that samples the population for characteristics such as income, housing, and education. After the 2000 Census, the long form was replaced by the American Community Survey (ACS). All three are discussed below.

With a few exceptions, the Census-derived time series presented in these profiles represent an amalgamation of data points from these three sources. While we made every effort to ensure comparability between figures over time, in some cases – detailed in table 2 – this was not possible and/or was difficult to assess. Furthermore, for the sake of narrative efficiency, we indicated all ACS data as corresponding to 2010 throughout the text and charts, even though the majority of it actually corresponds to the five-year timeframe between 2005 and 2009.

Please note that, for tabulation purposes, the Census treats cities as political units rather than spatially-fixed communities. As such, apparent changes over time may reflect changes caused by annexation, as well as changes within the original city boundaries. The table below indicates the extent of annexation for each of the ten case cities between 1970 and 2010.

Table 1. Change in land area by city, 1970–2010

City	Land Area in Square Miles		Percent Change
	1970	2010	
Fort Wayne	51.5	110.6	115%
Gary	42.0	49.9	19%
Grand Rapids	44.9	44.4	-1%
Pontiac	19.7	20.0	1%
Aurora	14.1	44.9	219%
Joliet	16.5	62.1	276%
Racine	13.1	15.5	18%
Green Bay	41.7	45.5	9%
Cedar Rapids	50.7	70.8	40%
Waterloo	59.2	61.4	4%

Notes: 1. Data for 1970 come from 1972 County and City Databook as accessed through ICPSR.
2. Data for 2010 come from the U.S. Census Bureau State and County Quickfacts.

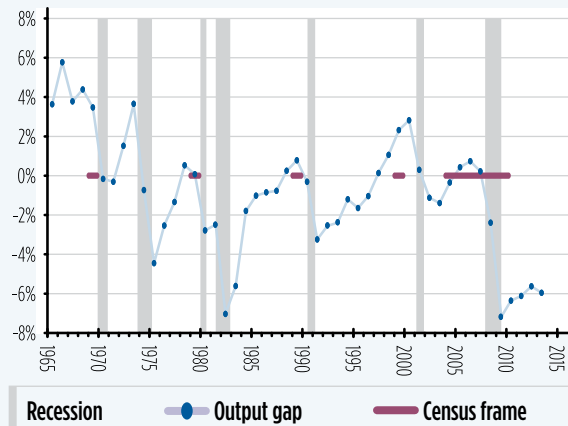## Inset 1: Census data and the business cycle

For most characteristics, observed changes over time neatly capture the long-term trends that interest us. For a handful of characteristics, however, historically meaningful structural changes may be somewhat obscured by short-term fluctuations in the business cycle. To illustrate, Census data indicate that real median family income in Green Bay increased by just over 12 percent between 1990 and 2000. This probably understates the true gain, however, insofar as the first measurement reflects income closer to the peak of a business cycle than the second one.[1]

This concern mainly applies to income- and employment-related characteristics. Ideally, in the interest of holding cyclical change constant and thereby isolating structural change, comparisons between these types of characteristics should be made between measurements taken during the same stage of the business cycle (e.g., peak-to-peak or trough-to-trough). When not possible, however, such comparisons should at least take into account that differences in timing with respect to the business cycle may be relevant.

Chart 1. Real U.S. output gap as a percent of real potential GDP

Source: Congressional Budget Office/Haver Analytics.

These differences are captured in chart 1, which displays the timeframe for income questions (Census frame) from the Census and ACS in relation to fluctuations in the business cycle. Note that both the formal definition of business cycles (in shading, and an informal measure depicted by the output gap (i.e., the difference between actual GDP and potential GDP), are depicted. The output gap rises during economic expansions and falls during contractions. We express it as a percent of real potential GDP to isolate this cyclical effect from long-term, structural increases in GDP. In the context of our example, the red line in 1989 highlights the period for which income was reported in the 1990 Census and the red line in 1999 highlights the same for the 2000 Census. Visually, we can see that the 1990 frame is closer to a recession and decline in the output gap; indicating it occured closer to the peak of a business cycle.

Lastly, in addition to the official U.S. Census website for sharing recent data (American FactFinder), for historical data we relied on two intermediary venues that organize the myriad older Census products into a coherent framework. In particular, for the period 1970-1990, we relied heavily on the National Historical Geographic Information System (NHGIS) maintained by the University of Minnesota. As a supplement, we also used data provided by the Interuniversity Consortium for Political and Social Research (ICPSR) maintained by the University of Michigan. Accordingly, the full citation for any specific Census-derived figure should be considered as "[the source] as obtained through [the venue], [the year]". Additional detail for each of these venues is provided below.

Sources

[i] Short Form

Citation: *U.S. Census Bureau, Decennial Census, Short Form.*

In contrast to the long form or ACS, all persons complete the short form. All households and group quarters receive a questionnaire by mail every ten years. It asks for the age, sex, and race/ethnicity for each person living at the address, as well as whether the residence is owned or rented.[2] Addresses are primarily obtained from the Master Address File from previous Census years and the Delivery Sequence File from the U.S. Postal Service.[3] Follow-ups are conducted by telephone and personal interviews for nonrespondents. Missing data are imputed. Since the published figures are enumerations and not estimates from a sample, there are no calculable margins of error associated with sampling bias. However, the decennial Census is accompanied by a post-enumeration survey to assess coverage error.[4] The post-enumeration survey for the 2010 Census did not find a significant percent net undercount or overcount for the household population.[5]

[ii] Long Form

Citation: *U.S. Census Bureau, Decennial Census, Long Form.*

For Censuses 1970-2000, one in six residents received a long form questionnaire with detailed questions on population and housing. Though results from the long form are technically estimates (not enumerations), the Census Bureau considers the figures sufficiently precise that it does not publish margins of error.

[iii] American Community Survey

Citation: *U.S. Census Bureau, American Community Survey.*

The Census Bureau officially introduced the ACS in 2005 as a replacement for the Decennial Census long form. Instead of sampling the population at one point in time every ten years, the ACS draws monthly rolling samples from U.S. households and group quarters for release every year.[6] Because these annual samples are smaller than the long form samples (about 1 in 40), geographies with smaller populations require greater than single-year periods to achieve appropriate margins of error.[7] Thus the ACS also releases rolling three-year and five-year estimates, where the multi-year estimates are constructed by pooling data from all years. For our analysis of industrial cities, appropriate margins of error were typically only obtainable from 5-year data. In some cases, our assessment of the standard error relative to the estimate allowed us to use three-year data (this measure is known as the coefficient of variation (CV); see discussion below for additional detail). It should be noted that we only considered margins of error when selecting the timeframe for an estimate. We did not test whether differences in estimates are statistically significant. Comparisons of ACS data made in the profiles may not be statistically significant when the estimates are very close or from a small population.

[iv] County and City Data Book

Citation: *U.S. Census Bureau, County and City Data Book [United States] consolidated files, 1944-1977.*

The County and City Data Book is a compendium of local-area data compiled by the U.S. Census Bureau from a variety of sources. It was published as a supplement to the Statistical Abstract of the United States in 1952, 1956, 1962, 1972, 1977, 1983, 1988, 1994, 2000, and 2007.[8] For budget reasons, the Bureau terminated the program in 2011.

Venues

[i] American Factfinder

Citation: *U.S. Census Bureau, American FactFinder, http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.*

American FactFinder provides access to data about the United States, Puerto Rico, and the Island Areas. The data in American FactFinder come from several censuses and surveys.

For more information see "Using FactFinder" and "What We Provide."[9, 10]

[ii] NHGIS

Citation: *Minnesota Population Center. National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota 2011, http://www.nhgis.org.*

The National Historical Geographic Information System (NHGIS) provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2012.

[iii] ICPSR

Citation: *The Interuniversity Consortium for Political and Social Research. Ann Arbor, MI: University of Michigan, http://www.icpsr.umich.edu/.*

The Interuniversity Consortium for Political and Social Research maintains an extensive archive of data sets in the social sciences. Data are available to researchers at no charge.

[iv] Miscellaneous

Percent manufacturing in 1960 and two other national figures for 1970 were not found in the above venues and thus obtained elsewhere, as indicated below.

- Percent Manufacturing from University of Virginia Library
  Citation: *University of Virginia Library, County and City Data Books, http://www2.lib.virginia.edu/ccdb.*

- Median Family Income from Current Population Reports
  Citation: *U.S. Census Bureau, U.S. Department of Commerce, Current Population Reports, Consumer Income, Series P-60, No. 78. May 20, 1971, http://www2.census.gov/prod2/popscan/p60-078.pdf.*

- Median Value of Owner Occupied Homes from Historical Census of Housing Tables
  Citation: *U.S. Census Bureau, U.S. Department of Commerce, Historical Census of Housing Tables, Home Values, http://www.census.gov/hhes/www/housing/census/historic/values.html.*

Table 2. U.S. Census figures by Decennial Form

Order	Figure	Description	Census Form	Notes
1	Total population	Total number of persons	Short	--
2	% < 19	% of total population aged 19 and under	Short	--
3	% 20-24	% of total population aged 20-24	Short	--
4	% 25-44	% of total population aged 25-44	Short	--
5	% 45-64	% of total population aged 45-64	Short	--
6	% > 65	% of total population aged 65 and over	Short	--
7	% Black	% of population that identified themselves as Black	Short	To ensure comparability with earlier years, universe is constrained to persons who identified with only one race.
8	% White	% of population that identified themselves as White	Short	To ensure comparability with earlier years, universe is constrained to persons who identified with only one race.
9	% Hispanic or Latino (of any race)	% of total population that reported a Hispanic country of origin	Short	Not found for 1970 and 1980. Unlike race figures, universe includes the entire population.
10	% Less than HS	% of population aged 25 and over that did not graduate from high school	Long	See % HS Grad note.
11	% HS Grad	% of population over 25 who graduated from high school but never attended college	Long	In 1970, there is no explicit distinction between high school graduate and non-high school graduate. Individuals assumed to have graduated high school if and only if they completed 4 years of high school.
12	% Some College & College Grad	% of persons aged 25 and over that ever attended college	Long	--
13	% Manufacturing	% of employed population aged 16 and over that work in the manufacturing industry	Long	Figures for 1970 appear to omit approximately 3-8% of eligible universe. Figures for 1960 come from County and City Data Book.
14	Civilian Work Force	Full civilian work force, including the unemployed	Long	--
15	% Civilian Unemployed	% of individuals who are in the labor force but not employed	Long	--
16	Real Median Family Income	Real median family income, adjusted using CPI-U-RS (2010=100)	Long	See extended note to figure 16 below.
17	% Families Below Poverty Line	% families below poverty line	Long	--
18	Mean Commute Time	Mean travel time to work (minutes)	Long	Only found for 2000 and 2010.
19	% Married (individuals 15 years and over)	% of population aged 15 and over that are married	Long	In 1970, includes persons 14 years and over.
20	Average HH size	Average number of persons per household	Short	Only found for 2000 and 2010.
21	Average Family Size	Average family size	Short	Not found for 1970 and 1980.
22	Total Units	Total number of housing units	Short	--
23	% Owner Occupied	% of occupied housing units that are owner occupied	Short	--
24	Real Median Value of Owner Occupied Homes	Real median value of specified owner occupied homes	Long	See extended note to figure 24 below.
25	% homes w- 0 Vehicle	% of occupied units with no vehicles	Long	--
26	% homes w- 1 Vehicle	% of occupied units with exactly 1 vehicle	Long	--
27	% homes w- 2+ Vehicles	% of occupied units with 2 or more vehicles	Long	--

Table 2. U.S. Census Figures by Decennial Form

28	% Foreign Born	% of entire population that was born abroad to non-native parents	Long	See extended note to figure 28 below.
29	Real Median Household Income	Real median household income, adjusted using CPI-U-RS (2010=100)	Long	See extended note to figure 29 below.
30	% Rent Burden	% of renting HHs whose gross rent is greater than or equal to 35% of income	Long	See extended note to figure 30 below.

General notes

In all cases:

- All data from 2000 and after were obtained through American FactFinder.

- Non-ACS figures that take into account income (median family income, median household income, and rent burden) are based on income from the year immediately prior to the indicated year (e.g., 1970 income data corresponds to 1969); the timeframe for ACS income-related figures is also offset by one year (e.g., income data from the 2005-2009 timeframe corresponds to 2004-2008).

- Real dollar amounts were adjusted using the CPI-U Research Series (CPI-U-RS, 2010=100).

Unless otherwise indicated:

- Figures indicated as deriving from the "Short Form," do in fact derive from the Decennial Census Short Form for all years.

- Figures indicated as deriving from the "Long Form" derive from the Decennial Census Long From for all years except 2010; in that case, data were derived from the 2005-2009 American Community Survey.

- All figures from 1960-1990 were obtained through the NHGIS.

Extended notes to figures

16   In 1970, city- and state-level figures were taken from the County and City Data Book as obtained through the ICPSR, while the U.S. level figure was taken from a Current Population Reports publication (see http://www2.census.gov/prod2/popscan/p60-078.pdf). We were unable to find sufficient documentation to confirm comparability between 1970 and later years.

24   The following caveat applies to comparisons between 1970 and later years: For 1980-2010, the population of units includes only "specified" units, which represents a subset of single-family homes (see http://quickfacts.census.gov/qfd/meta/long_HSG495210.htm for the definition of "specified" as employed in the ACS). In 1970, however, city- and state-level figures were taken from the County and City Data Book as obtained through the ICPSR. The codebook entry for that year is indicated as "OOU.SINGLE FAMILY MEDIAN VAL. $1970." We were unable to determine if this contains all single family homes, or just a subset thereof. The U.S. level figure for 1970 was obtained from Historical Census of Housing Tables (see http://www.census.gov/hhes/www/housing/census/historic/values. html), and appears to subset the population of units in a manner consistent with the definition of "specified." Any potential difference in the underlying universe should be mitigated by our using the median rather than the mean.

28   For 1970 and 2000: We assume, but cannot verify, that "foreign" excludes individuals born abroad to native parents. In Joliet in 1970, 2.3% of the eligible universe appears to be missing. For the last data point, we used a narrower three-year timeframe (2009-2011), as the coefficients of variation were generally acceptable. The CV for Gary, however, straddled the informal threshold between "Good" and "Fair".

29   We assume, but cannot verify, that the population includes all households, as opposed to a subset of households that meet a certain criteria. For 2010, we used ACS data from the 2009-2011, as all coefficients met the informal criteria for "good" reliability.

30   2010 figures correspond to ACS five-year estimates from the 2007-2011 timeframe. Due to changes in the universe, comparability might be problematic for 1970, and is definitely problematic for 2007-2011. Figures relating to 1980-2000 all take into account "specified renter occupied housing units," while 1970 takes into account "renter-occupied units for which rent tabulated," and 2010 takes into account "renter-occupied housing units." The Census Bureau makes the disclaimer that the ACS data is not suitable for comparison with earlier long form data due to this change in the universe. By this logic, 1970 may be problematic as well. Renters who did not pay rent or who had a non-positive income are omitted from all calculations. Although we cannot verify the definition of gross rent for all years, in recent years "Gross rent is the contract rent plus the estimated average monthly cost of utilities...and fuels...if these are paid for by the renter." (For example, see http://www.socialexplorer.com/data/ACS2012/metadata/?ds=Social+Explorer+Tables%3A++ACS+2012 +(1-Year+Estimates)&table=T102B.)

Inset 2: Detailed discussion of ACS reliability and the coefficient of variation

Inherent in the design of the ACS is a tradeoff between timeliness, accuracy, and geographic specificity; given limited resources and therefore a limited sample size, it's impossible to have all three of these desirable properties simultaneously.

To give researchers better control over how exactly these tradeoffs are calibrated, the ACS provides estimates of demographic characteristics in terms of 5-year, 3-year, and 1-year timeframes. The 5-year estimates are the most reliable because they have the largest sample size. Furthermore, 5-year estimates are available for all geographies for which the ACS tabulates data. The obvious downside of the 5-year data is that it applies to a long period, and may therefore be unsuitable for understanding short-term trends and/or the current picture. The 1-year data, on the other hand, is suitable for analyzing short-term dynamics. The downside is that it is only available for larger geographies, and that estimates may have a high margin of error. The properties of the 3-year data are somewhere in between those of the 1-year and 5-year data.

Given that we are dealing with midsize cities, the choice was really between the 3-year and 5-year estimates. (1-year estimates are available for most cities, but omit Pontiac as well as several cities used for comparison. Further, as will be explained below, cities that barely met the population thresholds for inclusion in the 1-year data may suffer from high margins of error that would make their use questionable.)[11]

To make the decision between the 3-year and 5-year data, we follow the Census Bureau's advice and look at a metric known as the Coefficient of Variation (CV). The Bureau emphasizes that an acceptable CV should ultimately be a function of the estimate's intended use, and declines to provide specific interpretive thresholds. However, an informative user guide compiled by the Washington State Office of Financial Management suggests that, as a general rule, estimates with CVs less than 15% may be considered "good," estimates with CVs between 15% and 30% may be considered "fair," and estimates with CVs in excess of 30% should be used "with caution."[12]

Throughout, we only used 3-year data when the CVs were acceptable for all case study cities.

[2] U.S. Bureau of Labor Statistics

[i] Quarterly Census of Employment and Wages

Citation: *Bureau of Labor Statistics, U.S. Department of Labor, Quarterly Census of Employment and Wages [www.bls.gov/cew/].*

Employment and location quotient data by industry are from the Quarterly Census of Employment and Wages as obtained through the Location Quotient Calculator.[13] Employment is calculated from quarterly reports filed by nearly every employer in the U.S.[14]

When used in the profiles, these data reflect annual averages for the county corresponding to the case-study cities. Please see below for the definition of "location quotient." Information on living wage calculations, which generally accompany these data in the profiles, is provided in A-9.

[ii] Occupational Employment Statistics

Citation: *Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics, (www.bls.gov/oes/).*

Employment, location quotient, and wage data by occupation are from the May 2012 release of the Occupational Employment Statistics for Metropolitan and Nonmetropolitan Areas. These estimates were calculated based on a rolling sample of establishments from May 2012, November 2011, May 2011, November 2010, May 2010, and November 2009.[15] The Employer Cost Index is used to express wage data across the timeframe in terms of May 2012 constant dollars.

When used in the profiles, these data reflect figures for the CBSA or Metropolitan Division corresponding to the case study cities. Please see below for the definition of "location quotient." Information on living wage calculations, which generally accompany these data in the profiles, is provided in A-9.

[iii] Employment Projections

Citation: *Bureau of Labor Statistics, U.S. Department of Labor, Employment Projections (www.bls.gov/emp/).*

All employment and output projections by industry are at the national level, and were taken from table 2.7 of the 2010-2020 Employment Projections Program.[16]

Inset 3: Location Quotient Definition

A location quotient (LQ) measures the concentration of a characteristic in one level of geography relative to that same concentration in a reference geography.[17] In the profiles, we employ location quotient to examine employment by industry between county and U.S., and employment by occupation between MSA and U.S.

LQs greater than one indicate that the characteristic is more concentrated in the local geography than the nation, while LQs less than one indicate it is less concentrated. For example, the 2011 LQ of paper manufacturing in Kane County, IL, is 2.43. This means that the share of paper manufacturing employment in Kane County is 2.43 times greater than the national share.

Mathematically, a LQ is a representation ratio defined by:

$$LQ = \frac{e_i / e}{E_i / E}$$

Where:

e_i = Local employment in industry i

e = Total local employment

E_i = Base area employment in industry i

E = Total base area employment

[3] CPI-U-RS

Citation

- For 1978 and onward: U.S. Bureau of Labor Statistics, Consumer Price Index Research Series Using Current Methods (CPI-U-RS), U.S. city average, all items, December 1977=100 (see http://www.bls.gov/cpi/cpiursai1978_2012.pdf).

- For years prior to 1978: extrapolations as calculated by the U.S. Census Bureau (see http://www.census.gov/hhes/www/income/data/incpovhlth/2012/CPI-U-RS-Index-2012.pdf).

All values presented in real dollars were adjusted for inflation using the Consumer Price Index research series (CPI-U-RS) as employed by the U.S. Census Bureau. The CPI-U-RS is officially published by the Bureau of Labor Statistics (BLS) for a period beginning in 1978.[18] The Census Bureau derives values for prior years by applying the ratio of the CPI-U-RS and CPI-U in 1977 to the 1947-1976 CPI-U. Though the index is published such that December 1977=100, we transformed the series to present values in terms of 2010 dollars.

The CPI-U-RS tracks historical changes in the cost of living more consistently and accurately than the commonly reported Consumer Price Index for All Urban Consumers (CPI-U). It is more consistent because it applies current methodology to all years in the series, while the CPI-U – despite improving over the years – is not adjusted retroactively. Incorporating these improvements, in turn, improves accuracy. Current methods have reduced upward bias, which the Boskin commission reported to be 1.1 percent per year.[19] For example, the CPI now accounts for lower-level substitution bias (i.e., substitutions made among purchases within the same class of good.) Accordingly, the research series exhibits lower rates of inflation than the CPI-U. These improvements are especially significant for longitudinal analysis where rates compound over time. The CPI-U estimates that the price level rose by 462 percent between 1970 and 2010, whereas the CPI-U-RS estimates the increase at 401 percent.[20]

It should be noted that the CPI-U-RS, while an improvement over the CPI-U, still does not represent the BLS' best measure of a cost-of-living index because it does not accommodate for substitutions made between classes of goods (aka, upper-level substitutions).[21] To appreciate the significance of this type of substitution, it's helpful to note that a cost-of-living index should estimate the increase in income necessary to make a consumer just as happy after an increase in the price level as before. As an example, if the price of pork increases relative to beef, a consumer may be just as happy purchasing more beef and less pork. Thus an index which presumes the consumer purchases the same amount of pork at a higher price is upwardly biased. The BLS produces a series that accounts for this effect, the Chained CPI-U, but it only extends back to year 2000.[22] Examining the change in price level between 2000 and 2010 (years for which all three indices are available), the Chained CPI estimates an increase of 23 percent, while the CPI-U and CPI-U-RS both estimate an increase of 27 percent.[23]

It should also be noted that the CPI-U-RS is a national index and may not reflect regional differences in the cost of living across the 10 cities. Thus readers are cautioned against interpreting cities with comparatively lower median incomes or median incomes that fail to keep pace with the CPI-U-RS as strictly worse off.

[4] HMDA

Main Citation: *Federal Financial Institutions Examination Council (FFIEC), Home Mortgage Disclosure Act (HMDA) loan application register flat files (http://www.ffiec.gov/hmda/hmdaflat.htm).*

Tract-to-City Crosswalk: *2000 U.S. Census Bureau boundary data, as obtained through Maptitude Version 5.*

The Home Mortgage Disclosure Act (HMDA) requires that certain lending institutions publically report information pertaining to loan applications for home purchases, improvements, and refinancing.[24] Policymakers and regulators use the resulting report – which includes borrower characteristics such as race and income – to assess whether institutions are meeting the credit needs of the community, as well as to deter discriminatory practices. In addition to these regulatory purposes, the data are well suited to place-based analysis in general because they include the Census tract of the property.

In the profiles, we limited our data to home purchase loans that were either originated or denied by the lending institution after a full review of the application. Preapprovals and withdrawn applications were not considered. Data were aggregated by Census tract and then converted to city-level data using 2000 Census boundary data as obtained through Maptitude. All dollar values were adjusted for inflation using the CPI-U-RS.

[5] CRA

Main Citation: *Federal Financial Institutions Examination Council (FFIEC), Community Reinvestment Act (CRA) aggregate flat files (http://www.ffiec.gov/cra/craflatfiles.htm).*

Tract-to-City Crosswalk: *2000 U.S. Census Bureau boundary data, as obtained through Maptitude Version 5.*

The Community Reinvestment Act (CRA) requires certain depository institutions to report data on business lending for the public.[25]

Data include loans made in amounts of less than $1 million; to better focus on lending to small businesses we further limit the data to loans made to businesses with less than $1 million in revenues. Tract-level data was converted to city-level data using 2000 Census boundary data as obtained through Maptitude. All dollar values were adjusted for inflation using the CPI-U-RS. Note that, unlike HMDA, CRA does not provide data regarding applications.

[6] FDIC Summary of Deposits

Main Citation: *FDIC Summary of Deposits (http://www2.fdic.gov/sod/).*

Geocoding-related Citations:

- Maptitude Version 5.

- 2000 U.S. Census Bureau boundary data, as obtained through Maptitude Version 5.

- The Google Geocoding API, Version 2 (https://developers.google.com/maps/documentation/geocoding/).

- Federal Reserve Bank of Chicago calculations.

The Federal Deposit Insurance Corporation (FDIC) Summary of Deposits is an annual report that reflects, among other things, the geographic distribution of deposits held by all FDIC-insured institutions. Information in the report is obtained from two sources: 1) a mandatory survey required of all FDIC-insured institutions that operate two or more branch locations, including foreign institutions that operate in the U.S. and 2) the Call Report, which may be used in place of the survey in cases where an institution operates in only one location.[26] These data comprise the vast majority of deposits and deposit-like instruments held in the U.S.; credit unions – whose deposits collectively summed to about 12 percent of that of commercial banks in 2004 account for the remainder.[27]

In the survey, institutional respondents are asked to allocate total deposits to physical bank locations in a manner consistent with their respective internal practices.[28] For example, the allocation of a certain account to a certain branch office for SOD purposes might derive from matching the account holder's address to the nearest branch, where the account is most active, or where the account was opened.

Furthermore, respondents are instructed to consolidate the deposits of limited-service outlets (such as ATMs) into more substantial branches located nearby (preferably in the same county). The sum of deposits distributed over the various locations should match the analogous figure in the Call Report or Report of Assets and Liabilities.[29]

The subsequent availability of detailed address fields in the report can be used to pinpoint the exact latitude and longitude of bank locations (and their corresponding deposits), thereby making this source particularly useful for the sort of place-based analysis employed throughout the profiles. This process of converting addresses to coordinates is known as "geocoding", and is implemented by a piece of software called a "geocoder."

We used two geocoders to match deposits with the profiled cities: Maptitude (v5) and the Google Geocoding API (v2). After determining the coordinates of bank locations, we then used Maptitude again to determine the corresponding city with respect to boundaries from the 2000 Census.

It is important to note that all geocoders rely on matching techniques with degrees of uncertainty in order to reconcile text-based address fields between multiple data sources. Consequently, any geocoding procedure is subject to multiple types of error including: 1) failure to match at all, 2) matching to the wrong location, and 3) matching to a correct but imprecisely defined location (e.g., a zipcode as opposed to a building).

Regarding the first type of error, our geocoding success rate generally fell between about 90 percent and 95 percent, depending on the year. The second type of error, while important, is difficult to quantify. Since our goal was to link banking data with a relatively large target (cities), we imagine that the third type of error is insignificant.

A few general caveats are worth mentioning given how deposits are reported and geocoded:

- First, note that deposits figures reported throughout the profiles relate to deposits corresponding to bank locations in the cities, not residents of the cities. Throughout the profiles, however, we implicitly presume that these two measures are highly correlated, and use them interchangeably.

- Second, between the survey instructions and Banks' internal practices, an area's figures may be skewed upward if it contains a central location within which large amounts of deposits from nearby limited-service locations are consolidated. (This effect was particularly noticeable in the case of Green Bay, WI, where one location with consolidated deposits drove per-capita deposits to a level nearly three times higher than that of the next highest case study city.)

- Lastly, given that geocoding outcomes tend to be more successful for recent periods than for earlier periods, estimated growth in deposits may be subject to upward bias. Using two geocoders mitigates but does not eliminate this bias.

Miscellaneous notes:

- While all discussions pertaining to deposits amounts draw from geocoded data, discussions relating to institutional characteristics and market structure (e.g., number of branches, market share, community versus non-community bank) draw from Summary of Deposits data as assigned to cities based on their zipcodes. This assignment, in turn, was based on 2000 city and 2007 zipcode boundaries from the Census, as obtained through Maptitude.

- The FDIC began including the results of its internal geocoding procedure starting with the 6-2012 release. All deposits figures in our analysis, however, are entirely based on geocodes obtained through Maptitude and Google as described above.

- Data were aggregated by Census tract and then converted to city-level data using 2000 Census boundary data as obtained through Maptitude. All dollar values were adjusted for inflation using the CPI-U-RS.

[7] LPS Applied Analytics

Main Citation: *Lender Processing Services (LPS) Applied Analytics.*

Zipcode-to-City Crosswalk: *2000 U.S. Census Bureau boundary data, as obtained through Maptitude Version 5.*

Proprietary loan-level microdata furnished by LPS Applied Analytics details the monthly performance of mortgage loans in the residential housing market. LPS collects this data from large mortgage servicers, who collectively represent about two-thirds of this market.

The underlying raw data include numerous mortgage types including first mortgages, second mortgages, and various grades of home equity lines of credit. In an effort to better align our measures with properties as opposed to loans, however, we take into account only first-lien mortgages. Furthermore, we used Census data (as obtained through Maptitude V5) to assign loans to case study cities using the zipcode of the underlying property.

A variety of possible metrics may be derived from mortgage performance data to help gain insight into the health of a given housing market, including but not limited to: the foreclosure start, transition, and inventory rates. Throughout the profiles, we focus exclusively on the foreclosure inventory rate, a static measure that represents the number of mortgages in foreclosure as a proportion of all mortgages. The start and transition rates, on the other hand, are dynamic measures that provide insight into the flow of loans into and out of foreclosure status.[30]

It's important to note that foreclosure inventory rates are highly sensitive to state laws that govern how foreclosures are processed. A foreclosure in Illinois, for example, takes about 300 days and often longer because every foreclosure must be processed through the courts. However, some states, like Michigan, do not require foreclosures to go through the courts. Still, depending on the situation, certain states like Iowa and Wisconsin employ both methods. All things being equal, foreclosure rates tend to be lower in states that rely primarily on non-judicial procedures, as any potential buildup resulting from new foreclosures in these states is tempered by the speed with which they can be resolved.[31]

Given this sensitivity to various legal procedures, foreclosure inventory rates should only be compared among states with similar process periods. In the profiles, we compare the foreclosure inventory rate in a given city with its home state and the average of a group of reference states. The four reference groups were constructed based on the quartiles of the process period, as shown in table 3.

Table 3. Typical foreclosure process period for reference states

Group	Process Period (days)	States
1	< 63	AL CT DC GA MD **MI** MO NH RI TN TX VA WY
2	63-136	AK AR AZ CA FL KS MA MN MS NC NV VT WA WV
3	136-180	CO **IA** ID KY LA MT ND NE NM OR SC SD UT
4	>180	DE HI **IL IN** ME NJ NY OH OK PA **WI**

Source: RealtyTrac (see http://www.realtytrac.com/real-estate-guides/foreclosure-laws/).

[8] Brown University

Citation: *Spatial Structures in the Social Sciences, Brown University, US2010 Project, (http://www.s4.brown.edu/ us2010/Data/data.htm).*

Measures of residential segregation and racial/ethnic composition are from US2010, a project of Spatial Structures in the Social Sciences at Brown University, and based on data from the Decennial Census and the 2005-09 American Community Survey.

The dissimilarity index measures the extent to which one group is distributed proportionally across census tracts in a city relative to another group.[32] The index ranges from 0 to 100 and equals zero if every tract exhibits the same ratio between groups as the city as a whole. The index equals 100 if the two groups are entirely segregated by census tract. Values of 60 or above are considered fairly high. It means that 60 percent of one group must move to a different tract to achieve a proportional distribution. Values between 40 and 60 are considered moderate, while values less than 40 are fairly low.

More generally, the index for two racial groups is defined as:[33]

$$\frac{1}{2} \sum_{i=1}^{N} \left| \frac{x_i}{X} - \frac{y_i}{Y} \right|$$

Where:

x_i = the population of group X in census tract i

X = the total population of group X in the city

y_i = the population of group Y in census tract i

Y = the total population of group Y in the city

[9] Living Wage Project

Citation: *Poverty in America, Massachusetts Institute of Technology, Living Wage Project, Living Wage Calculator (http://livingwage.mit.edu/).*

Estimates of living wages are from the Living Wage Calculator, a tool provided by the Living Wage Project under the Poverty in America program at the Massachusetts Institute of Technology. A living wage represents a minimum cost of living for low wage families in a particular area based on cost estimates for food, child care, healthcare, housing, transportation, other necessities, and taxes. It is intended to highlight that working families may not earn enough to live locally, even if they earn more than the minimum wage and are not officially in poverty.

All estimates cited in the profiles are for one adult raising one child. The calculator uses data from a variety of federal sources to estimate costs, including the Bureau of Labor Statistics, the U.S. Department of Housing and Urban Development, and the U.S. Department of Agriculture. Estimates are made with respect to the latest source data that was available in June 2012.

Though the calculator allows users to select estimates for either place or county, it does not detail the various levels of geography represented by the source data. Therefore we cannot distinguish which cost estimates, if any, are particular to the place or county, and which represent some broader level of geography. Estimates cited in the profiles were selected by place, and these are likely more representative of the MSA or metropolitan division, where one exists.

Additionally, the calculator does not report whether values are given in constant dollars. Given the latest update in June 2012, we speculate that all values can be generally assumed to be in "recent" dollars.

Notes

1. As the table below indicates, please note that income reported in the 1980 and 1990 Census corresponds to income from 1979 and 1989, respectively.

2. U.S. Census Bureau, Explore the Form, available at http://www.census.gov/2010census/about/interactive-form.php.

3. U.S. Census Bureau, Summary Population and Housing Characteristics, Selected Appendixes, May 2012, available at http://www.census.gov/prod/cen2010/cph-1-a.pdf.

4. U.S. Census Bureau, Coverage Measurement, available at https://www.census.gov/coverage_measurement/.

5. U.S. Census Bureau, Census Coverage Estimation Report, May 2012, available at http://www.census.gov/coverage_measurement/pdfs/g01.pdf.

6. U.S. Census Bureau, American Community Survey, Design and Methodology, available at http://www.census.gov/acs/www/methodology/methodology_main/.

7. Basic information on sample size and data quality by state can be found at http://www.census.gov/acs/www/methodology/sample_size_and_data_quality/.

8. U.S. Census Bureau, County and City Data Book: 2007, available at http://www.census.gov/prod/2008pubs/07ccdb/ccdb-07.pdf.

9. U.S. Census Bureau, Using FactFinder, available at http://factfinder2.census.gov/faces/nav/jsf/pages/using_factfinder.xhtml.

10. U.S. Census Bureau, What We Provide, available at http://factfinder2.census.gov/faces/nav/jsf/pages/what_we_provide.xhtml.

11. U.S. Census Bureau, American Community Survey, Guidance for Data Users, available at http://www.census.gov/acs/www/guidance_for_data_users/estimates/.

12. Washington State Office of Financial Management, American Community Survey User Guide, May 2012, available at http://www.ofm.wa.gov/pop/acs/userguide/ofm_acs_user_guide.pdf.

13. Bureau of Labor Statistics, Quarterly Census of Employment and Wages, Location Quotient Calculator, available at http://data.bls.gov/location_quotient/ControllerServlet.

14. Bureau of Labor Statistics, Quarterly Census of Employment and Wages, Frequently Asked Questions, available at http://www.bls.gov/cew/cewfaq.htm#Q14.

15. Bureau of Labor Statistics, Occupational Employment Statistics, Overview, available at http://www.bls.gov/oes/oes_emp.htm.

16. Bureau of Labor Statistics, Employment Projections, available at http://bls.gov/emp/ep_table_207.htm.

17. Bureau of Labor Statistics, Help & Tutorials, available at http://www.bls.gov/help/def/lq.htm#location_quotient.

18. Bureau of Labor Statistics, CPI Research Series Using Current Methods, available at http://www.bls.gov/cpi/cpirsdc.htm.

19. Bureau of Labor Statistics, Price Measurement in the United States: a decade after the Boskin Report, Monthly Labor Review, May 2006, available at http://www.bls.gov/opub/mlr/2006/05/art2full.pdf.

20. Calculated from the annual averages of the national CPI-U, All items as obtained from http://www.bls.gov/cpi/data.htm.

21. Bureau of Labor Statistics, Frequently Asked Questions about the Chained Consumer Price Index for All Urban Consumers, available at http://www.bls.gov/cpi/cpisupqa.htm

22. Bureau of Labor Statistics, Note on the Chained Consumer Price Index for All Urban Consumers, available at http://www.bls.gov/cpi/superlink.htm.

23. Calculated from the annual averages of the national Chained CPI-U, All items as obtained from http://www.bls.gov/cpi/data.htm.

24. Depository and non-depository institutions alike are covered by HMDA, subject to their asset size, presence in the MSA, and whether they are involved in the business of residential mortgage lending. See page 3 of the HMDA reporting guide (http://www.ffiec.gov/hmda/pdf/2010guide.pdf) for details.

25. Subject to asset thresholds updated annually (for example, see: http://www.ffiec.gov/cra/pdf/Explanation%20of%20the%20Community%20Reinvestment%20Act%20Asset%20Threshold%20Change%20121712.pdf), all state member banks, state nonmember banks, national banks, and savings associations are required to report. Institutions that do not meet these thresholds have the option of reporting voluntarily.

26. Federal Deposit Insurance Corporation, Summary of Deposits Reporting Instructions, available at http://www2.fdic.gov/sod/pdf/SOD_Instructions.pdf, page 1.

27. Federal Reserve Bank of San Francisco, Are credit unions regulated or supervised by the Federal Reserve System?, Dr. Econ blog, March 2005, available at http://www.frbsf.org/education/publications/doctor-econ/2005/march/credit-unions-regulation-supervision.

28. Federal Deposit Insurance Corporation, Summary of Deposits Reporting Instructions, available at http://www2.fdic.gov/sod/pdf/SOD_Instructions.pdf, page 1

29. Ibid, page 3.

30. For a detailed discussion of how these rates interrelate, please see our guest blog at http://midwest.chicagofedblogs.org/archives/2011/10/emily_engel_for.html.

31. Lower inventories, however, do not necessarily translate into healthier housing markets. Properties that moved through foreclosure quickly in Michigan, for example, may show up subsequently as real estate owned (REO) by the mortgagee. We do not track post-foreclosure statuses like REO because we're unsure to what extent LPS tracks them.

32. Spatial Structures in the Social Sciences, Brown University US2010 Project, Interpreting a Data Set, available at http://www.s4.brown.edu/us2010/Data/Explanation.htm.

33. Population Studies Center, University of Michigan, Racial Residential Segregation Measurement Project, available at http://enceladus.isr.umich.edu/race/calculate.html.

INDUSTRIAL CITIES INITIATIVE

FEDERAL RESERVE BANK
OF CHICAGO

COMMUNITY DEVELOPMENT AND POLICY STUDIES