LONGITUDINAL EMPLOYER–HOUSEHOLD DYNAMICS

TECHNICAL PAPER NO. TP-2007-03

Social, Economic, Spatial, and Commuting Patterns of Self-Employed Jobholders

Date: April 2007
Prepared by: Paul Ong & Matthew R. Graham
Contact: U.S. Census Bureau
4600 Silver Hill Rd.
LEHD Program
Attn: Jeremy S. Wu, Rm 6H141
Suitland, MD 20746 USA

This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. This document is released to inform interested parties of research and to encourage discussion. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854, and the Alfred P. Sloan Foundation. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see <http://www.ces.census.gov/>). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, 4600 Silver Hill Rd., LEHD Program, Data Integration Division, Rm 6H141, Suitland, MD 20746, USA. (Jeremy.S.Wu@census.gov <http://lehd.did.census.gov/>).

Social, Economic, Spatial, and Commuting Patterns of Self-Employed Jobholders*

April 2007

Paul Ong**
School of Public Affairs
UCLA

Matthew R. Graham
Abt Associates and
U.S. Census Bureau

---

* The authors wish to acknowledge the LEHD program and its staff for making this data available and providing support for these analyses. We would also like to thank Lucy Tran for her assistance in assembling the background materials, working on the graphics, and editing of this report.

** Corresponding Author: pmong@ucla.edu
School of Public Affairs, UCLA

Abstract
A significant number of employees within the United States identify themselves as self-employed, and they are distinct from the larger group identified as private jobholders. While socioeconomic and spatial information on these individuals is readily available in standard datasets, such as the 2000 Decennial Census Long Form, it is possible to gain further information on their wage earnings by using data from administrative wage records. This study takes advantage of firm-based data from Unemployment Insurance administrative wage records linked with the Census Bureau's household-based data in order to examine self-employed jobholders – both as a whole and as subgroups defined according to their earned wage status – by their demographic characteristics as well as their economic, commuting, and spatial location outcomes. Additionally, this report evaluates whether self-employed jobholders and the defined subgroups should be included explicitly in future labor-workforce analyses and transportation modeling. The analyses in this report use the sample of self-employed workers who lived in Los Angeles County, California.

Keywords
Self-employed jobholders; labor-workforce; commuting patterns; modal split; transportation modeling; OD-matrix; Los Angeles; California; administrative wage records; U.S. Census Bureau; Longitudinal Employer-Household Dynamics; LEHD.

Abbreviations

ACS – American Community Survey
CPS – Current Population Survey
JH – Jobholder
LEHD – Longitudinal Employer-Household Dynamics
LF – 2000 Decennial Census Long Form, Average 1-in-6 Sample
POR – Place of Residence
POW – Place of Work
PIK – Protected Identification Key
PMSA – Primary Metropolitan Statistical Area
SBO – Survey of Business Owners

INTRODUCTION

This is the third of three technical briefs examining workers who are difficult to study because of the unique nature of their labor market status: multiple jobholders, informal jobholders, and self-employed jobholders. This report covers the analysis of self-employed jobholders, individuals who reported being self-employed but might have recorded wage earnings in employment data during the first two quarters of 2000. The analyses in these reports provide new insights by combining data from two sources: the 2000 Census and the LEHD Program, both of which are described later. Each dataset has strengths (the Census data is population-based while the LEHD data is firm-based), and when combined, they provide complementing coverage and data items. The analyses have a number of objectives, but one of the most important is to examine the commute characteristics of self-employed workers who were unpaid (according to administrative wage records) relative to the commute characteristics of self-employed workers who received paid wages and who constitute a minority of the total sample. While there are some studies that examine the personal characteristics of these workers and their employment outcomes, there is a paucity of information on the commute itself: travel mode, time, and distance, as well as origins and destinations. These characteristics are critical both to the study of the spatial structure of metropolitan labor markets and to transportation analysis and modeling.

The analysis in this report focuses on self-employed jobholders and compares the two subgroups of the self-employed that are based upon whether workers were paid and had their wage information reported to the unemployment insurance (UI) system.[*] This data does not currently contain information on self-employed workers who do not contribute to state UI programs, and thus any analysis using only UI data misses socioeconomic and commuting data from this class of people. By combining LEHD's UI data[†] with 2000 Census Long Form (LF) data, we are able to get a better measurement of those who are self-employed as well as more information about their economic and transportation characteristics. Throughout this report the region of study is

[*] Self-employed jobholders are defined herein as workers who self-reported in the 2000 Census that they were self-employed during the enumeration period. These workers were broken into two groups based upon whether they had recorded wage earnings in either Q1 or Q2 of 2000 according to LEHD data.

[†] It is important to note that while "UI data" and "LEHD data" might be used interchangeably here, the actual UI data is only a portion of the entire data available to LEHD.

limited to the Los Angeles-Long Beach Primary Metropolitan Statistical Area (PMSA).[*]

In this analysis of self-employed jobholders there are three key questions that should be resolved: (1) Are self-employed workers adequately captured by LEHD data? (2) Is the self-employed jobholder population (or its constituent subgroups) sufficiently large that it cannot be ignored in labor-workforce analyses and transportation modeling? and (3) Are the personal, employment, and commuting characteristics of this population and its subgroups sufficiently distinct from each other and from the privately employed population that they need to be explicitly factored into labor-workforce analyses and transportation modeling? The first question is immediately answered from the fact that 14.0% of the non-government, civilian workforce living in Los Angeles County were classified as self-employed, with over two-thirds of these workers not showing up in the official wage earning records.[†] This number of self-employed workers, which is in the range of previously reported data reviewed in the next section, suggests that LEHD is not adequately capturing data for self-employed workers, or at least the subgroup that does not contribute directly to unemployment insurance programs. Again, the size of the self-employed jobholder pool suggests that this group of workers is of significant importance for future analyses of the metropolitan labor market and modeling of the commute-to-work patterns. The overall commute and spatial patterns among paid and unpaid self-employed jobholders are distinct, and there are some important differences that suggest that self-employed jobholders should be explicitly included in future transportation analyses and modeling.

BACKGROUND

The reported size of the self-employment sector varies with the data source. The Decennial Census provides a conservative count. Of the 129.7 million employed civilian population aged 16 and over reported in the 2000 Census, 8.6 million or 6.6% were self-employed workers (U.S. Census Bureau, DP-3), a category that by definition includes “people who worked for profit or fees in their own unincorporated business, profession, or trade, or who operated a farm.” Another

[*] The Los Angeles-Long Beach Primary Metropolitan Statistical Area (PMSA) and Los Angeles County are coterminous and lie completely within the five-county Los Angeles-Riverside-Orange County Consolidated Metropolitan Statistical Area (CMSA).

[†] The estimates in this report are based on responses from a sample of the population. As with all surveys, estimates may vary from the actual values because of sampling variation or other factors. All statements made in this report have undergone statistical testing and are significant at the 90-percent confidence level, unless otherwise noted.

4.1 million or 3.2% were self-employed people who were paid employees of their own incorporated businesses (U.S. Census Bureau, QT-P25). If unpaid people who worked at least 15 hours in a business or on a farm operated by a family member are included, then the self employed sector represents 10.1% of the economically active workforce, or about 13.1 million individuals, with a large majority not appearing in administrative records on paid workers. However, because the Decennial Census only allows for one response for the question on "class of worker," the reported statistics do not include those who report being self-employed but who also work for pay outside their businesses.

An alternative source of self-employment data is the Survey of Business Owners (SBO), which provides statistics on the composition of U.S. businesses owned by individuals. The SBO includes all firms that reported $1,000 or more in receipts on their 2002 tax forms as individual proprietorships, partnerships, and any type of corporation. According to the SBO, there were almost 22.5 million privately held firms in 2002, with the vast majority of these businesses (77% or 17.3 million) not having any paid employees (U.S. Census Bureau, 2005). Many of the privately held firms (about 2.5 million) had equal ownership shared between males and females, and a fair number probably relied upon unpaid family members; consequently, the total number of individuals involved in the self-employed sector was probably close to 20.4 million individuals based on the SBO. This estimate is over one and a half times the number that the Census reports as being self-employed, which is consistent with the SBO finding that only two-thirds of business owners stated that their businesses were the primary source of income. In other words, the self-employed sector represents nearly a sixth of the economically active population according to the SBO.

Over time, the rate of self employment has fluctuated, as indicated by data from the Current Population Survey (Hipple, 2004). From 1976 to 2004, the number of unincorporated businesses increased from about 5.8 million to 9.5 million, with short-term movements that appear to be related to the business cycle. Despite this growth, the unincorporated self-employed as a percent of the labor force peaked in 1983 at 7.8%, up from 6.8% in 1976. By 2004, the percentage stood at 6.9%, essentially the same as at the beginning of this time period. However, the decline in the share of the labor force comprised by unincorporated self-employed workers was offset by an

increase in the share comprised by incorporated self-employed. Overall, the CPS data indicate that the self-employed sector has accounted for roughly a tenth of the active labor force and has contributed to the overall growth rate in total employment (Manser and Picot, 1999). Of course, the CPS data undercounts the number of self-employed for the same reasons that the Decennial Census does, including unreported self-employment in addition to a primary job.[*]

There are noticeable variations in the self-employment rate across demographic groups. This sector is disproportionately overrepresented by males, whites, and older individuals (Hipple, 2004). Although there has been considerable attention given to self-employment among immigrants, their share of the self-employed is actually lower than their share of wage and salary workers. There are minor differences in the self-employment rate by educational attainment, with a slight overrepresentation by those with the least and most education. Self-employment rates are particularly high among older workers, with 17.8% of the workers between the ages of 55 to 64 and 26.9% of those 65 or older in 2003 (Hipple, 2004; Karoly and Zissimopoulos, 2004).

There is also considerable demographic variation in the performance of firms owned by individuals, as indicated by 2002 SBO data (Ong, 2005). Women-owned businesses accounted for a little more than a third of privately-held firms, but their share of revenues and payroll is less than one sixth. The average of receipts for female-owned businesses was only a quarter of that for male-owned businesses. Blacks, American Indians and Alaskan Natives (AIAN), Native Hawaiians and other Pacific Islanders (NHOPI), and Hispanics were not only less likely to own businesses, but they were more likely to operate smaller businesses with limited receipts and payroll relative to non-Hispanic whites. In these respects, Asians are more comparable to the overall averages than other minority groups, but some disparity remains in terms of receipts and payroll.

Little is available in published form on the commute patterns of the self-employed. The U.S. Bureau of Labor Statistics does provide information on persons who usually worked at home based on the CPS (defined as "worked at home at least once per week as part of their primary

[*] A more detailed comparison between LEHD data with the CPS data would shed light on some questions raised here, but that analysis currently is outside the scope of this report.

job"). In 2004, roughly 15.1% of the employed population 16 and over fell into this category (U.S. Census Bureau, Table 2, 2005). Interestingly, the self-employed (both incorporated and unincorporated) comprised about a third of those who usually worked at home. This is three times higher than their share of all workers, indicating that about half of the self-employed conducted at least a part of their business at home. Moreover, this suggests that the commute patterns of the self-employed differ from those of workers with employers.

DATA, SAMPLE, AND METHODS

The Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program works with partner states to gather quarterly Unemployment Insurance (UI) data and develop data products for the states' use based upon that data. These UI data are collected by the states from employers and the datasets thus provide data primarily on jobs rather than on individuals. However, each worker is tagged with a Protected Identification Key (PIK) representing one jobholder and thus can be linked to other Census datasets following appropriate confidentiality processing so that socioeconomic factors can be matched with the quarterly UI earnings records.

The socioeconomic Census data used in this analysis came from the 2000 Decennial Census Long Form (LF) dataset, which is an average 1-in-6 sample of the entire U.S. population. Each record represents one respondent from the Decennial Census and contains socioeconomic data as well as commuting and geographical data on the individuals' places of residence and work. It also includes a weighting factor that allows the full population to be imputed.

By using LEHD data matched to Census LF data, we were able to differentiate between those self-employed jobholders who reported wages through the unemployment insurance (UI) system and those who did not and then to develop some insights into the characteristics of their jobs by industry and earnings, as well as places of work and residence.[*] Since in the strictest sense unpaid workers should not appear in LEHD data, we used the LF data and the lack of matching records in LEHD to identify this class of self-employed workers. Because we included those

[*] Throughout the remainder of this report, we will use the phrase "unpaid" to refer to those who reported themselves as being self-employed but earned no income during Q1 or Q2 of 2000 according to the UI data reported to LEHD and "paid" to refer to those who were self-employed and did have reported earnings in the UI data for the first half of 2000.

workers who had earnings in only one of the first two quarters in the "paid" category, this method tends to generate a conservative estimate of the unpaid self-employed jobholder population. Used as a baseline "control" were the privately employed population who identified having earnings in at least one of the first two quarters as reported through the UI system. For both paid and unpaid self-employed jobholders, the universe within the LF sample population was the same: individuals who identified themselves as currently self-employed either by incorporated firms or within unincorporated enterprises as well as being at-work during the enumeration period.

Four analyses were performed with the information generated by the dataset described above, and the Los Angeles metropolitan area (see footnote on page 5) was used as a case study for the four analyses. The first analysis examines and compares the social characteristics of unpaid self-employed jobholders and paid self-employed jobholders by using the merged data from the LF sample of the 2000 Census. By merging individual-level data from the 2000 Census, it is possible to examine the individual and household characteristics associated with being a self-employed jobholder. Specifically we considered sex, age, educational attainment, race/ethnicity, and nativity.

The second analysis focuses on employment outcomes, particularly sectoral (industrial) distribution and earnings. Total annual earnings from 1999, rather than quarterly earnings from LEHD, were used because they provide information on both those workers identified as unpaid self-employed as well as paid self-employed. Because this earnings data is from 1999 rather than 2000, it should be seen as the "earnings potential" for each worker in 2000.

The third analysis examines and compares the travel patterns of self-employed jobholders by using the LF data on an individual's commute to work, specifically travel mode and travel time. A comparison of distance between residence and work (or origin-destination) uses the LF's locational information. Because of uncertainty in the geocoding results of the locations – particularly the POWs – jobholders' residences and the firm locations were aggregated at the tract level and all distances were calculated between the tract centroids.

The fourth analysis examines and compares the spatial distribution of jobs held by paid self-employed jobholders and those held by unpaid self-employed jobholders by place of residence and place of work. Making use of the origin-destination matrix (by tract) constructed and summarized in the previous analysis, we compared the spatial distribution of PORs against POWs for each type of jobholder. Additional comparisons were performed between classes of POWs and PORs. Two analytical methods were used to compare the spatial distributions: the first was a simple correlation of tract origins and destinations, and the second was the calculation of dissimilarity indices, a method which is described below.

We use the dissimilarity index, a widely used measurement in the study of residential segregation (Iceland et al., 2002), as one method of comparing the spatial distribution of paid and unpaid self-employed workers. The index ranges from 0 to 100 and roughly indicates the percentage of a group that would have to move to achieve full integration across the universe of geographical units with another group. The following is an example of its application to the analysis in this technical report.

$$DI = 100 \cdot \sum_i \tfrac{1}{2} \left| \frac{p_i}{P} - \frac{q_i}{Q} \right|$$

where DI = Dissimilarity Index (0-100)

p_i = Count of first population in geographical unit i.

P = Count of first population across the universe of geographical units.

q_i = Count of second population in geographical unit i.

Q = Count of second population across the universe of geographical units.

In this report several different dissimilarity indices were calculated (e.g., between the PORs of paid self-employed jobholders and the PORs of unpaid self-employed jobholders). The geographical unit of analysis in this case is the Census Tract, and the universe of units is Los Angeles County.

FINDINGS

Overall, our sample of the self-employed represented 14.0% of the non-government workforce, and of this sample, 67.9% were unpaid and 32.1% were paid.* These figures are within the reported ranges of the literature and are comparable to expectations of the specific workforce composition of Los Angeles County. Of course, the sample does not include those workers in self-employed positions who were not enumerated by the Census or who held private employment as well as self-employment and reported only the private employment. Nor does it attempt to determine the size of the population that worked additional UI wage jobs at the same time as self-employment that did not produce official UI wages. The total sample size of self-employed jobholders was 389,396. This is approximately 18.8% of the jobholders who were identified in this project as "formal" workers (see Graham and Ong, 2005). This group of paid, privately employed jobholders is used throughout this report as a baseline for comparison of demographic, economic, commuting, and spatial trends.

Table 1: Demographic Characteristics

	Total Self-Emp Sample [%]	Unpaid Self-Employed [%]	Paid Self-Employed [%]	Formal Jobholders§ [%]
Sex				
Male	66.3	65.8	67.4	54.0
Female	33.7	34.2	32.6	46.0
Age				
16-24	2.67	2.0	4.1	15.4
25-44	45.0	42.5	50.0	55.3
45-64	44.6	46.4	40.9	27.4
65+	7.8	9.1	5.0	1.9
Education				
Less than HS	16.0	16.6	14.8	20.0
HS	14.6	15.0	13.6	18.8
Some College	29.0	29.2	28.5	33.7
BA/BS+	40.5	39.2	43.1	27.5
Race/Nativity				
U.S. Born	59.3	57.5	63.1	61.6
Naturalized	23.0	24.0	21.0	19.0
Non-citizen	17.7	18.5	16.0	19.5
Race/Ethnicity				
NH White	56.7	55.7	58.7	39.8
Black	4.5	4.4	4.8	7.8
Asian	14.2	14.6	13.4	13.4
Hispanic	19.8	20.4	18.7	33.6
Others	4.8	5.0	4.4	5.4
Incorporation Status				
Incorporated	27.5	18.0	47.5	NA
Not Incorporated	72.5	82.0	52.5	NA

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
§ For further information on the definition of "formal" jobholders, see Graham and Ong, 2005, which is the second report of this series.
Notes:
1. Total % for each group may not equal 100% due to rounding.
2. Certain values in this table may not be significantly different from one another.
3. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
4. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

Distinct from the ratio for overall jobholder trends, male self-employed jobholders outnumber females by nearly two to one (see Table 1). There is also a slightly larger discrepancy between males and females for paid self-employed jobholders (67.4% male vs. 32.6% female). In terms of age, self-employed jobholders are almost exclusively in the 25-64 age range, and again the

[*] Within the context of all three reports, it should be reiterated that the self-employed population is mutually exclusive from the "informal" population because informal workers self-report as "privately employed" instead of "self-employed," as the current sample does.

relation is even stronger among the subgroup identified as having paid earnings recorded by the UI system. One interesting note here is that unpaid self-employed jobholders tend to be older than paid self-employed jobholders suggesting that there may be some relationship between experience and the form of self-employment.

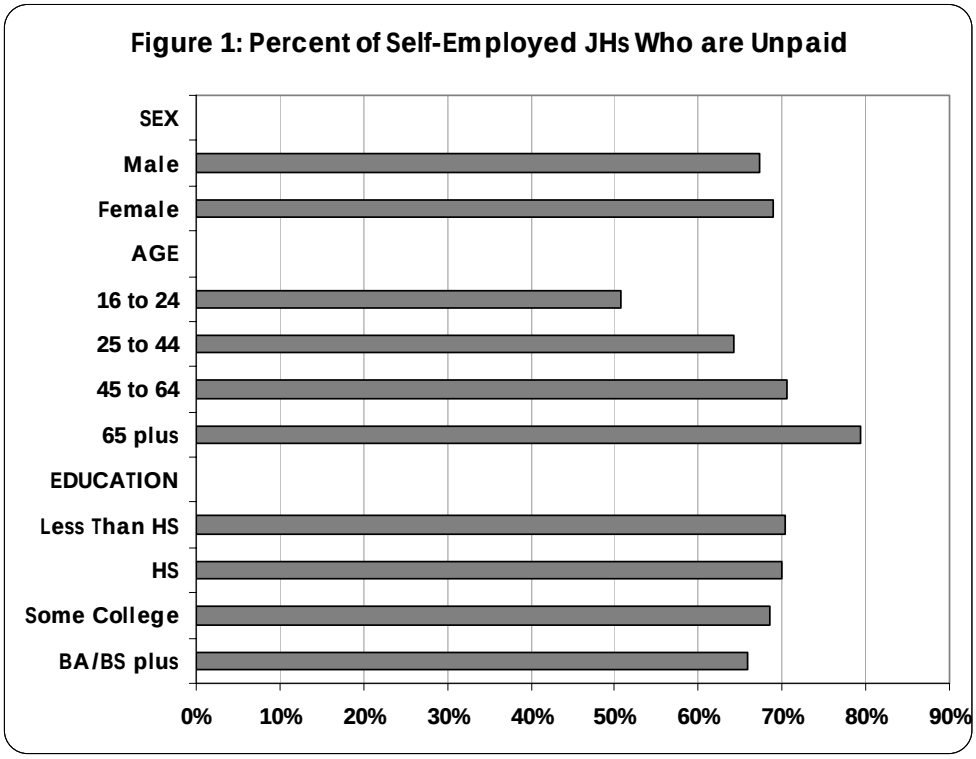
A similar but opposing trend appears for self-employed jobholders by educational attainment. Almost 70% of the self-employed have at least some college in their educational experience. But in this case, the trend is skewed toward the paid self-employed, 43.1% of whom had bachelor's degrees or better, while only 39.2% of the unpaid self-employed had similar educational experience. In terms of nativity, the U.S.-born population holds nearly 6 of every 10 self-employment positions, with naturalized citizens having a slight advantage over non-citizens for the remainder of those positions. Again, comparing between the paid and unpaid positions, U.S.-born jobholders have a 5.6% advantage for paid self-employment. By race/ethnicity whites hold the majority of self-employment positions and that portion is larger for paid than unpaid (58.7% of paid vs. 55.7% of unpaid).

When compared to the larger workforce, we find that self-employed jobholders are distinctly older than the baseline of formal jobholders. They are also much more likely to be NH white (and much less likely to be Hispanic). And these differences with the formal jobholders are even stronger for paid self-employed. By nativity, there are only slight differences between the formal jobholders and self-employed workers.

Also included among the reported demographic data are the percentages of individuals who identified themselves as being self-employed either within an incorporated firm or an unincorporated one. A few interesting observations on the self-employed populations can be made using these data because they serve to some extent as secondary checks on the LEHD paid/unpaid classifications. From the definitions of the paid and unpaid classes, one might expect that a great majority of the paid self-employed would work for incorporated firms and the same for unpaid self-employed in unincorporated enterprises, with very few in the other combinations. However, as can be seen in Table 1, this is obviously not the case. While 67.9% of our sample are classified as unpaid and 72.5% classify themselves as working for a firm that is not

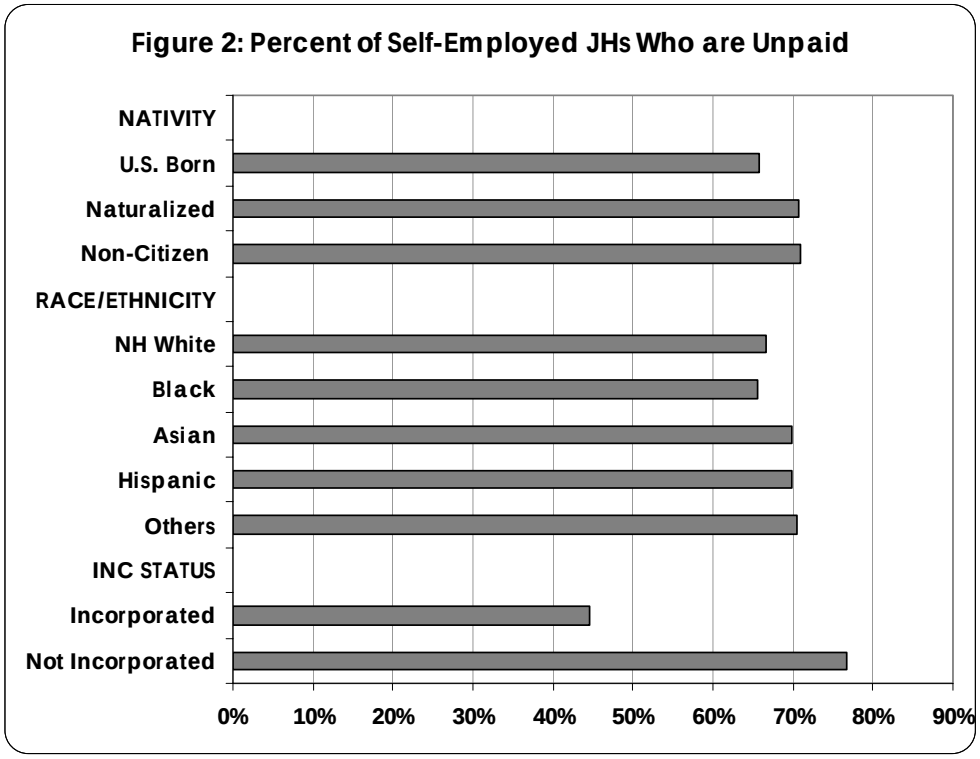
incorporated, the percentage breakdowns do fall into these two simple groups. For unpaid workers the percentages in incorporated (18.0%) and unincorporated (82.0%) firms approach what we might expect, though there are still almost a fifth of the workers who should appear in the administrative wage records but do not. When we turn to paid workers, we find they are not skewed so strongly in the opposite direction. In fact, with 47.5% of paid workers in unincorporated enterprises and 52.5% of them at incorporated firms, the unincorporated businesses still have the statistical advantage.

Another interesting point is that of those identifying themselves as working for incorporated firms, only 55.5% showed up in the official administrative wage records. This raises several potential problems. One, there is a question of whether these people are reporting in the long-form responses the correct legal status of their firm or whether these companies are properly reporting their data to the state UI system (at this point there is no information suggesting which is more important). Another possibility is that being incorporated does not automatically mean that the owner is a paid employee, and this has implications for how federal agencies (such as the Bureau of Labor Statistics) count the number of self-employed.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

Some of the relationships suggested above are illustrated explicitly in Figures 1 and 2. For example, unpaid self-employed jobholders have higher shares of each age category as they get older, with about 80% of the 65 and older category being unpaid compared to about 50% of the 16-24 category. By educational attainment, the relationship is obvious but not as strong: less education is frequently associated with higher rates of unpaid self-employment over paid self-employment.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

With respect to nativity, naturalized jobholders and non-citizens have similar rates of jobholding among the self-employed, and these are both approximately 5% higher than the rate for U.S.-born self-employed jobholders. By race/ethnicity, NH whites tend to hold less unpaid self-employment than do Asians, Hispanics, or other racial/ethnic groups. The differences here range between 3% and 5%. Finally, comparing incorporation status, about 75% of the unincorporated group were unpaid, while only approximately 45% of the incorporated group were unpaid.

Table 2 compares the distribution of jobs by industrial sector using the Census 2000 Industry Codes by 1-digit groupings. Relative to paid jobholders, unpaid jobholders maintain slight advantages in the fields of production, maintenance, installation, repair, office work, administration, management, business and financial services. The combined grouping of installation, maintenance, repair, and production (spread across two code groups) has the highest share of self-employment jobs with the next largest portion of the compositions being either

management, business, finance, and services for unpaid jobholders or farming, construction, and mining for paid jobholders.

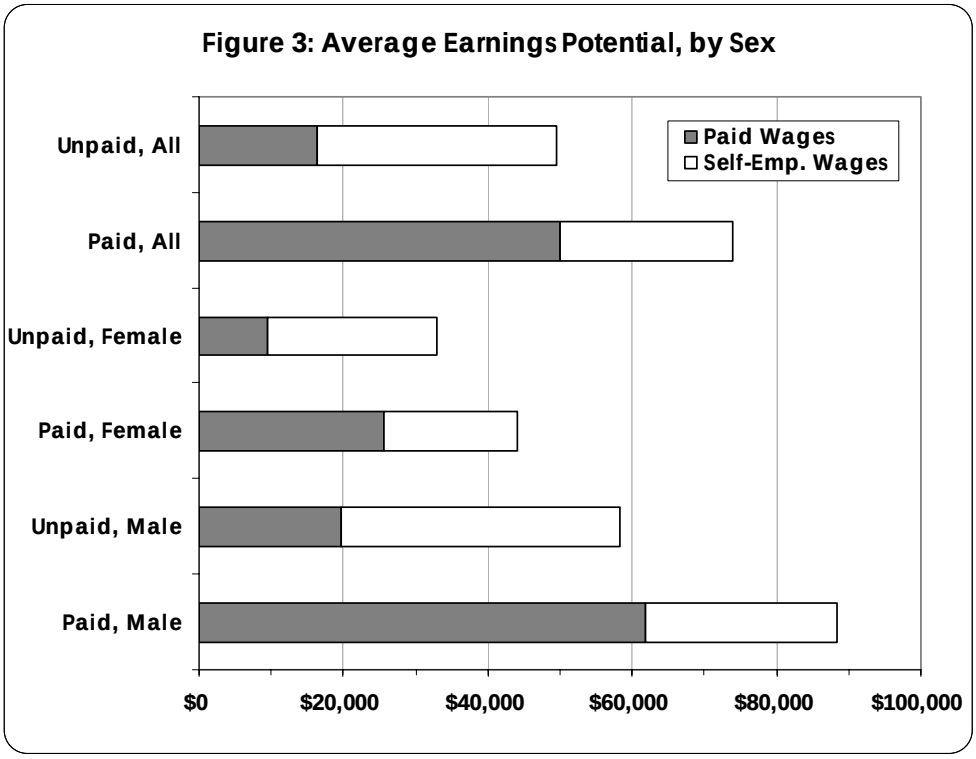
Table 2: Job Composition by Industry

	Unpaid [%]	Paid [%]
Mgmt., Bus., Fin. Svcs.	10.6	9.3
Comp., Arch., Eng., Sci.	2.0	2.9
Comm., Ed., Law, Arts, Sports, Media	0.8	1.6
Health, Protective Svcs.	2.1	3.6
Food, Sales, Maint., Pers. Svcs.	7.7	9.5
Office, Admin.	7.1	5.5
Farm, Constr., Mining	9.8	12.8
Install., Maint., Repair, Prod.	34.0	29.2
Add'l Prod.	21.9	21.6
Trans., Mil., Unempd.	4.0	4.1
Total	100.0	100.0

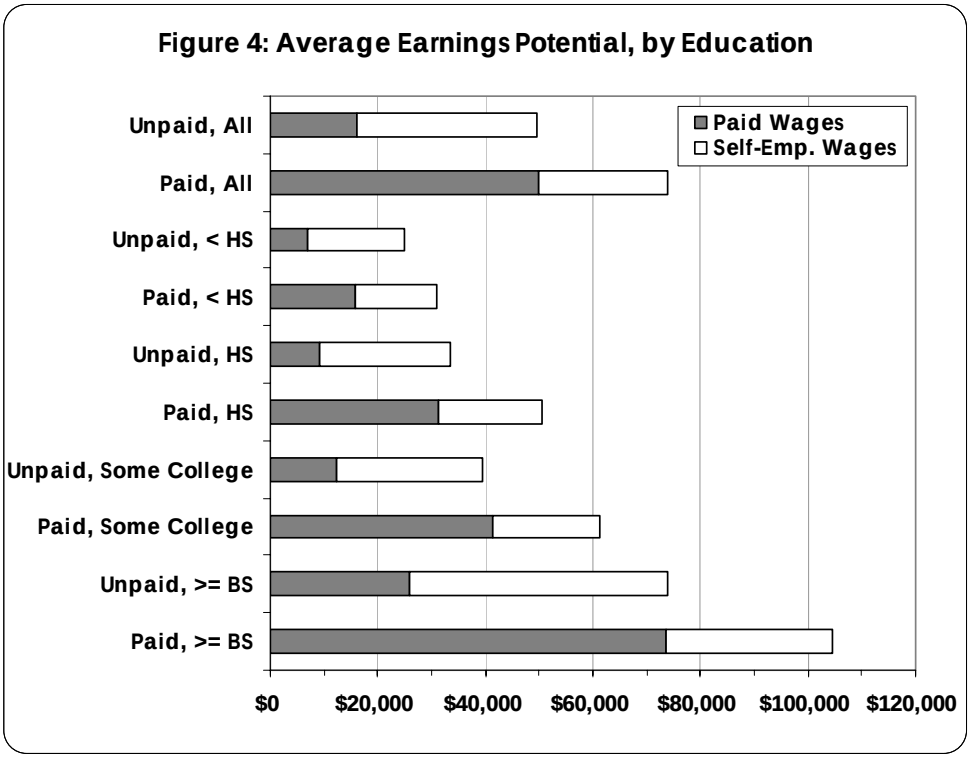
Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.



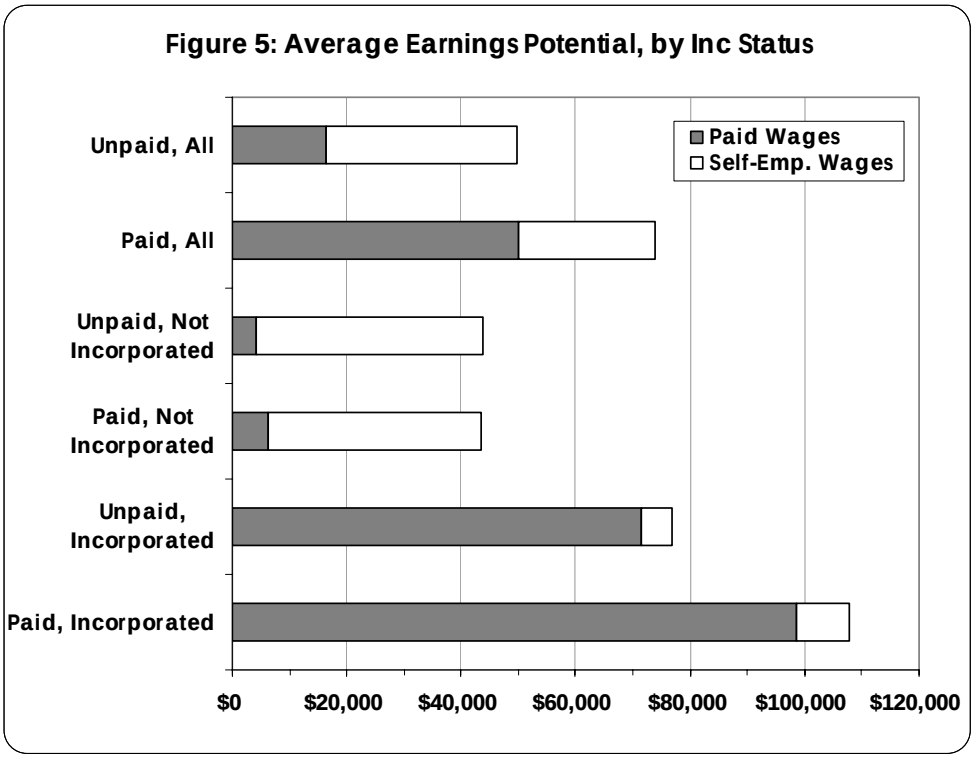
Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

Because earnings were, by definition, not available from UI data for unpaid self-employed workers, we used total reported 1999 wages from the LF data as a proxy for 2000 "earnings potential" in order to keep the earnings comparisons consistent between the self-employed subgroups. Also, the 1999 wages are reported in terms of both "paid wages" and "self-employed wages" as described by the Decennial Long Form Questionnaire. Even with this proxy variable, it is clear from the data (See Figures 3 and 4) that unpaid self-employed jobholders earn less – in some cases as much as 35% less – than their peers in terms of both sex as well as educational attainment. It is interesting to note, though, that on average men working at unpaid self-employed jobs earn more than women working at paid self-employed jobs, a fact that may reflect on different sectoral/pay-rate composition of jobs open to men and women in each class. The reporting of both paid and self-employed wages demonstrates explicitly that there is a distinct difference in the two subgroups as we have defined both in terms of whether they appear in UI data as well as how the individuals engaged in the different work activities classify their wages.

In every case, self-employed wages dominate the income of those groups we have defined as "unpaid," whereas paid wages dominate the incomes of those we defined as "paid." This is exactly as was expected.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

As can be seen clearly in Figure 5, the paid and unpaid self-employed wages are strongly associated with incorporation status. That is, for individuals who classified themselves as working for unincorporated firms the vast majority of their wages were self-employed wages, whereas for those who classified themselves as working for incorporated companies, they received mostly paid wages. However, because the relationship is still somewhat mixed, using the earnings ratio from the Census LF would not be exactly equivalent to the more detailed paid/unpaid classification used here. Also, it is immediately obvious that the incorporated self-employed earn significantly more than their unincorporated counterparts.

In Table 3 we can see that paid self-employed workers enjoy a distinct advantage over unpaid workers in the use of a car for commuting purposes, with 70.6% of paid workers driving alone compared to only 61.9% of unpaid workers. In turn unpaid workers are more likely to participate in a carpool or use "other" modes of transportation. Paid workers maintain a slight difference (0.6%) in use of public transit over unpaid workers.

Table 3: Commute Characteristics

	Unpaid [%]	Paid [%]	Formal§ [%]
Commute Mode			
Drive Alone	61.9	70.6	76.5
Carpool	13.2	11.8	14.0
Public Transit	1.7	2.3	4.9
Other	23.2	15.2	4.6
Commute Time			
0 to 14	41.1	32.7	20.3
15 to 29	26.1	29.7	34.7
30 to 59	24.7	28.9	33.9
60 plus	8.1	8.7	11.1
Commute Distance			
Same Tract	36.2	22.4	3.7
5 miles or less	26.3	28.1	33.3
5 to 15 miles	23.6	31.3	40.5
Over 15 miles	13.9	18.2	22.5

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
§ See Graham and Ong, 2005.
Notes:
1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

At the same time unpaid self-employed workers appear to spend less time commuting as well as having fewer long commute trips. These numbers are somewhat skewed by the significant difference in the numbers of self-employed jobholders working at home. For example, 20.1% of unpaid self-employed jobholders report their commute time as 0 minutes whereas only 12.4% of paid self-employed jobholders report that case. However, even after removing those self-employed jobholders with no commutes, the trend persists that the remaining unpaid workers still have shorter commutes than their paid counterparts. As a result, a similar skewing effect would also appear in the distance-to-work data, but the overall patterns still hold.

In a comparison with the baseline class of formal jobholders, the trends are obvious and distinct. Paid and unpaid self-employed jobholders have significantly shorter commutes both in terms of time and distance. This most likely leads to their lower rates of driving alone and use of carpools and public transit. The difference is made up in the "Other" category, which likely includes a large subgroup who work at home and do not commute at all.

Table 4 lists the various dissimilarity indices among places of residence (PORs) and places of work (POWs) for both the paid and unpaid self-employed jobholder classes within this study. It also includes dissimilarity indices constructed against the baseline formal jobholders, who were identified in the second report of this series. While not extremely low, these dissimilarity indices suggest that unpaid and paid jobholders' POWs as well as PORs are neither polarized nor are they perfectly distributed. In this case, PORs between the two subgroups are slightly better distributed among the tracts of Los Angeles County than are the POWs. As we might expect from the distance to work data in Table 3, POWs and PORs for unpaid self-employed workers are better distributed than those for paid workers (DI of 41.5 for unpaid vs. 49.4 for paid workers). This result coincides with the fact that paid workers tend to have longer commutes (and thus live in different tracts) than unpaid workers.

Table 4: Dissimilarity Indices Among Paid/Unpaid JH Classes

Group Comparison	Dissimilarity Index
Unpaid JH POW vs. Paid JH POW	39.1
Unpaid JH POR vs. Paid JH POR	36.0
Unpaid JH POW vs. Unpaid JH POR	41.5
Paid JH POW vs. Paid JH POR	49.4
Unpaid JH POW vs. Formal JH POW	44.8
Paid JH POW vs. Formal JH POW	31.6
All Self-Emp'd JH POW vs. Formal JH POW	32.6
Unpaid JH POR vs. Formal JH POR	35.8
Paid JH POR vs. Formal JH POR	37.2
All Self-Emp'd JH POR vs. Formal JH POR	25.3

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

When comparing to the baseline of formal jobholders, the dissimilarity indices are close to those between the different paid and unpaid self-employed groups. Overall, self-employed jobholders,

both as a single group and by their payment status, are moderately integrated spatially with the formal jobholders. For example, comparing the places of work for unpaid jobholders and formal jobholders, the DI is 44.8, which implies that somewhere between 40% and 50% of these populations would need to be redistributed to reach perfect integration. This is greater (less integrated) than the DI (39.1) for places of work between paid and unpaid self-employed. At the same time, the DI between unpaid self-employed jobholders and formal jobholders is only 31.6. Clearly, the spatial structure of residence and employment is complex and might easily create a situation where the jobs for paid and unpaid self-employed are more evenly distributed than those between the unpaid and formal jobholders but less so than between the paid and formal jobholders. Despite this complexity, the overall trend is for the self-employed (both in terms of jobs and residences) to be located somewhat differently than formal jobholders, though the discrepancies are not extreme.

Table 5: Correlation Coefficients among Paid/Unpaid JH Classes

	Paid Self-Empl'd JH POR	Unpaid Self-Emp'd JH POW	Paid Self-Emp'd JH POW
Unpaid Self-Emp'd JH POR	0.531 (<0.0001)	0.267 (<0.0001)	0.171 (<0.0001)
Paid Self-Emp'd JH POR		0.200 (<0.0001)	0.199 (<0.0001)
Unpaid Self-Emp'd JH POW			0.645 (<0.0001)

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.
4. Standard errors were calculated using SAS 9 and adjusted using the Census 2000 long form design factors.

As can be seen in Table 5, the correlation data for PORs and POWs aggregated at the tract level supports the DI calculations with high correlation coefficients between PORs of paid and unpaid self-employed workers as well as between POWs for the two classes of workers. In turn the coefficients between PORs and POWs are somewhat lower, which is unsurprising given the dissimilarity indices above. However, the coefficients are not negative, which also suggests that jobs and homes are not polarized, a condition which could result from the multipolar urban

structure of Los Angeles or the differing work and commute patterns that self-employed workers hold compared with private and government jobholders.

Table 6: Correlation Coefficients among Self-Employed/Formal JH Classes

	Formal JH POW	Formal JH POR
Unpaid Self-Emp'd	0.561	
JH POW	(<.0001)	
Paid Self- Emp'd	0.793	
JH POW	(<.0001)	
All Self- Emp'd	0.774	
JH POW	(<.0001)	
Unpaid Self-Emp'd		0.190
JH POR		(<.0001)
Paid Self- Emp'd		0.206
JH POR		(<.0001)
All Self- Emp'd		0.476
JH POR		(<.0001)

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.
Notes:
1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.
4. Standard errors were calculated using SAS 9 and adjusted using the Census 2000 long form design factors.

Table 6 presents the correlation coefficients between the self-employed groups and the group of formal jobholders, which is used as a baseline. Only the relevant coefficients have been given. As has been true throughout these analyses, places of work tend to be better correlated than places of residence, and in this comparison to the baseline, the difference is obvious. The POW for the group of all self-employed jobholders correlates very well with the formal jobholders' POWs, though the paid self-employed correlate much better than the unpaid self-employed. This case is quite different from the places of residence. Neither paid nor unpaid self-employed jobholders' PORs correlate particularly well with the PORs of formal jobholders. However, when the two groups are combined, the resultant group of all self-employed jobholders has a coefficient of 0.476 with formal jobholders' PORs.

CONCLUSION

After considering the results of this analysis, we must return to the three primary questions posited in the first section of this report: (1) Are self-employed jobholders adequately captured in

the LEHD data? (2) Is the pool of self-employed jobholders too large to be ignored in further analyses? and (3) Is the pool of self-employed jobholders too different to be ignored in further analyses? Also in question is whether the two self-employed subgroups identified here are different enough from each other to be considered separately in further labor-workforce and transportation analyses. The first question can be answered very quickly by returning to some of the initial statistics reported here. The self-employed population is significant in itself, but over two-thirds of that population do not report their wages through the official state UI mechanisms. Thus, it is clear that while some self-employed workers are being captured by the LEHD data, most are not. In addition, without the LEHD UI data, it would be impossible to make the classifications of paid and unpaid because the appropriate data is not available in the 2000 Long Form collection.

In addressing the second question, inspection of the relative size of the self-employed jobholder pool (14.0% of the non-government, civilian workforce) suggests this group is definitely of the size that would warrant the additional work and effort required to incorporate them explicitly into future analyses of the metropolitan labor market and into models of the commute-to-work patterns. As has been mentioned, our class of self-employed jobholders is a conservative estimate because of reporting issues for those holding private jobs as well as being self-employed, and so the share and importance of this group might well be greater than stated in this data. Even so, the percentage calculated within this report lies within reported ranges for informal workers and is comparable with estimates of informal workers in the Los Angeles area.

In terms of social characteristics, paid and unpaid classes of workers have some moderate differences, particularly in age, educational attainment, and nativity. Some differences also exist by race/ethnicity, and also by sex. However, for age and educational attainment the tendency is for older, less educated workers to hold unpaid jobs – a set of relations that may signify a split in the types of self-employed jobs being held. In terms of nativity, unpaid self-employed jobs tend to go to non-citizens over U.S.-born workers, but the relationships are not as clear in terms of race/ethnicity. The differences in industrial breakdown among the classes, though relatively small, are also as one might expect. However, when the job classes are compared by earnings

there are some clear and obvious differences, with paid workers having significantly higher earnings potential.

As was mentioned above, there are also some interesting discrepancies when we consider the group by incorporation status as reported in the LF data. While unpaid, self-employed workers have a distribution among incorporated and unincorporated firms that one might expect, the paid workers split almost 50/50. Thus, our expectation that those in an incorporated firm would mostly show up in the official UI data was not borne out. It is unclear at this point why these people do not show up in the administrative records, but it is certainly an issue worth further exploration.

When we compare the two groups by commuting patterns, we find that the differences remain, especially considering the rate of driving alone to work as well as the distance/time to work. Some of this can be explained by the relatively high rate of unpaid workers who have no commute, though it does not account for all of the difference between the two groups. And of course, when we compare the self-employed population with those who were privately employed (formal jobholders), significant differences appeared in a number of the demographic, economic, and commuting categories. The self-employed population was overrepresented by white males, and they tend to be older and more educated. In addition, the self-employed also have much shorter commute patterns, and consequently different commuting mode patterns, including less use of cars (driving alone) and carpools.

Finally, the patterns in spatial distribution are somewhat less distinct. Paid and unpaid jobholders tend to be more integrated (according to the calculations of dissimilarity indices) in where they live than where they work, but not by much, and these groupings are only slightly more integrated than places of work and residence. However, by correlations, there are more significant differences between the groups. POWs tend to correlate fairly well among the paid and unpaid self-employed as do the PORs, whereas the correlations are not so strong between PORs and POWs. The same differences in DIs and correlation coefficients exist when we compare self-employed workers to formal workers – there are lesser differences among dissimilarity indices than there are among the correlation coefficients. And while POWs are

significantly better correlated than PORs, the segregation results are mixed and depend on whether we consider the whole self-employed groups or one of the subgroups.

Overall, there are important differences in socioeconomic, commuting and spatial patterns – especially when comparing to the private sector employment – and self-employed workers should definitely be considered as a separate group for modeling and analysis purposes, which was the purpose of the third thematic question.

One significant caveat that must be mentioned is the likelihood that these results are strongly linked to the urban structure and economic spatial distribution of Los Angeles itself. As a multipolar city-region with complex commute patterns, these results may not be as applicable for those cities with a more classical polar/CBD (Central Business District) structure. Another factor germane to this report and worth further study in comparisons with other metropolitan areas is the effect of recent immigration and low-wage workers, which should be examined to a greater depth but which is beyond the scope of this project.

As for the analysis itself that was used for this study, the methodology of linking UI data and socioeconomic data from the Census LF appears to offer a rich resource in studying transportation among particular classes of workers. In addition, it appears that this methodology can be easily ported to analyses involving the American Community Survey, which has a smaller sample size but more current statistics than the Decennial Census.

ACCURACY OF THE ESTIMATES

The data contained in this report are based on the sample of households who responded to the Census 2000 long form. Nationally, approximately one out of every six housing units was included in this sample. As a result, the sample estimates may differ somewhat from the 100-percent figures that would have been obtained if all housing units, people within those housing units, and people living in group quarters had been enumerated using the same questionnaires, instructions, enumerators, and so forth. The sample estimates also differ from the values that would have been obtained from different samples of housing units, and hence of people living in those housing units, and people living in group quarters. The deviation of a sample estimate from the average of all possible samples is called the sampling error.

In addition to the variability that arises from the sampling procedures, both sample data and 100-percent data are subject to nonsampling error. Nonsampling error may be introduced during any of the various complex operations used to collect and process data. Such errors may include: not enumerating every household or every person in the population, failing to obtain all required information from the respondents, obtaining incorrect or inconsistent information, and recording information incorrectly. In addition, errors can occur during the field review of the enumerators' work, during clerical handling of the census questionnaires, or during the electronic processing of the questionnaires.

While it is impossible to completely eliminate error from an operation as large and complex as the decennial census, the Census Bureau attempts to control the sources of such error during the data collection and processing operations. The primary sources of error and the programs instituted to control error in Census 2000 are described in detail in Summary File 3 Technical Documentation under Chapter 8, "Accuracy of the Data," located at <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.

Nonsampling error may affect the data in two ways: (1) errors that are introduced randomly will increase the variability of the data and, therefore, should be reflected in the standard errors; and (2) errors that tend to be consistent in one direction will bias both sample and 100-percent data in that direction. For example, if respondents consistently tend to underreport their incomes, then

the resulting estimates of households or families by income category will tend to be understated for the higher income categories and overstated for the lower income categories. Such biases are not reflected in the standard errors.

All statements in this Working Paper have undergone statistical testing and all comparisons are significant at the 90-percent confidence level, unless otherwise noted. The estimates in tables, maps, and other figures may vary from actual values due to sampling and nonsampling errors. As a result, estimates in one category used to summarize statistics in the maps and figures may not be significantly different from estimates assigned to a different category. Standard errors for the Dissimilarity Index estimates were calculated using derived methods based upon folded normal distribution theory. See Elandt, 1961. Further information on the accuracy of the data is located at <http://www.census.gov/prod/cen2000/doc/sf3.pdf>. For further information on the computation and use of standard errors, contact the Decennial Statistical Studies Division at 301-763-4242.

REFERENCES

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2005. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." *LEHD Technical Paper*. No. TP-2006-01. Washington, D.C.: U.S. Census Bureau.

Elandt, R.C. "The Folded Normal Distribution: Two Methods of Estimating Parameters from Moments." *Technometrics*. 3(4): 551-562.

Graham, Matthew R., and Paul Ong. 2007. "Social, Economic, Spatial, and Commuting Patterns of Informal Jobholders." *LEHD Technical Paper.* No. TP-2007-02. Washington, D.C.: U.S. Census Bureau.

Hipple, Steven. 2004. "Self-employment in the United States: An Update." *Monthly Labor Review* 127(7): 13-23.

Iceland, John, Daniel H. Weinberg, and Erika Steinmetz. 2002. "Racial and Ethnic Residential Segregation in the United States: 1980-2000." *U.S. Census Bureau Series CENSR-3*. Washington, D.C.: U.S. Government Printing Office.

Karoly, Lynn A., and Julie Zissimopoulos. 2004. "Self-employment among Older U.S. Workers." *Monthly Labor Review* 127(7): 24-47.

Manser, Marilyn E., and Garnett Picot. 1999. "The Role of Self-employment in U.S. and Canadian Job Growth." *Monthly Labor Review* 122(4): 10-25.

Ong, Paul. 2005. *The Policy Challenge: Growing Women- and Minority-Owned Businesses*. Los Angeles: Ralph and Goldy Lewis Center for Regional Policy Studies.

U.S. Census Bureau. "United States Census 2000: Summary File 3 (SF 3) – Sample Data." *American FactFinder*. Accessed 16 October 2005 <http://factfinder.census.gov>.

U.S. Census Bureau. "DP-3. Profile of Selected Economic Characteristics: 2000." *American FactFinder*. Accessed 16 October 2005 <http://factfinder.census.gov/servlet/QTTable?_bm=y&-geo_id=D&-qr_name=DEC_2000_SF3_U_DP3&-ds_name=D&-_lang=en&-redoLog=false>.

U.S. Census Bureau. "QT-P25. Class of Worker by Sex, Place of Work, and Veteran Status: 2000." *American Factfinder*. Accessed 22 October 2005 <http://factfinder.census.gov/servlet/QTTable?-geo_id=04000US40&-qr_name=DEC_2000_SF3_U_QTP25&-ds_name=DEC_2000_SF3_U>.

U.S. Census Bureau. "Advance Report on Characteristics of Employer Business Owners: 2002 Business Provides Primary Source of Personal Income for the Owner." *2002 Survey of Business Owners (SBO).* Accessed 16 October 2005 <http://www.census.gov/econ/census02/sbo/income.htm>.

U.S. Census Bureau. 2004. "Industry, Occupation, & Class of Worker." American Community Survey. Accessed 16 October 2005 <http://148.129.75.16/acs/www/UseData/Def/Io.htm>.

U.S. Census Bureau. 2005. "2002 Survey of Business Owners Preliminary Estimates of Business Ownership by Gender, Hispanic or Latino Origin, and Race: 2002." *2002 Survey of Business Owners (SBO)*. Revised 28 July 2005 <http://www.census.gov/csd/sbo/>.

U.S. Census Bureau. 2005. "Table 2. Job-related Work at Home on Primary Job by Sex, Marital Status, Presence and Age of Children, Class of Worker, and Pay Status, May 2004." *Current Population Survey (CPS)*. Accessed 22 October 2005 <http://www.bls.gov/news.release/homey.t02.htm>.