# CHAPTER 16

# A METHOD FOR DETERMINING THE SIZE OF THE UNDERGROUND CASH ECONOMY FOR COMMERCIAL SEX IN SEVEN US CITIES

BILAL KHAN, MITCH DOWNEY, MEREDITH DANK, AND KIRK DOMBROWSKI

In this chapter, we seek to (1) derive a more rigorous estimate of the underground commercial sex economy (UCSE) in seven major US cities and (2) provide an understanding of the structure of this underground economy.[1] To estimate the size of the UCSE accurately, we had to produce simultaneous estimates in each city of the size of the cash-based trade in both illegal drugs and illegal firearms. Ordinarily, estimates for each of these are independent of one another and often for a single locale at a single time. Our approach goes in the opposite direction. The operating assumption of the estimation process that follows is that estimates of the size of various domains of the underground economy (UE) are more accurate when comparative data across time and across different locations are taken into account and when estimates of the size of

[1] This research was supported by Award No. 2010-IJ-CX-1674 from the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, conclusions, and recommendations expressed in this document are those of the authors and do not necessarily reflect those of the Department of Justice or of the Urban Institute, its trustees, or its funders. Elements of this chapter were reported to the National Institute of Justice in the form of a final technical report as per grant obligations. The authors would like to thank (1) final technical report coauthors Cybele Kotonias, Debbie Mayer, Colleen Owens, Laura Pacifici, and Lilly Yu and (2) the National Institute of Justice and Janine Zweig, senior fellow at the Urban Institute's Justice Policy Center, for their careful review of these findings.

one domain of the UE are forced to balance estimates of other domains in the UE with which they coincide.

Underground economies are pervasive, but the cash trade in commercial sex, illegal guns, and illegal drugs are concentrated and most readily observable in urban settings. As discussed here, including multiple urban sites in a single estimation process provided the opportunity to leverage *relative* estimates of the size of each UE domain across cities as elements in a large system of linear constraints. Here we take up the size of illegal gun, commercial sex, and illegal drug economies in Atlanta, Denver, Dallas, Miami, San Diego, Seattle, and Washington DC.[2] Further, by making use of multiple proxy measures within each UE domain (illegal gun, commercial sex, and drugs), we mitigate the influence of errors in any single proxy and distribute the dependence of the final estimates for all cities across a wide range of measured sources.

In seeking to estimate the relative sizes of UE domains, arrest data are problematic as a proxy measure, as law-enforcement strategies vary considerably across US cities, even more so when evaluated across time. Using arrest data as the basis for proxy measurement of various UE domains across a wide range of cities requires making the erroneous assumption of uniform arrest rates across geographical location and time. Fortunately, viable non-arrest-based data are available for use as proxy measures for both the illegal drug and illegal gun trades. As few non-arrest-based sources of data are available for estimation of the UCSE, we used interviews with arrestees to estimate the size and income levels of pimps/traffickers, in addition to sex workers, in each city to collect the data necessary to complete this estimation process. Because these data came from arrestee sources, special means had to be developed to compensate for the data-collection source. These corrections draw from what is normally considered a major challenge to single-market size specification: the fact that UCSE participants (especially pimps/traffickers) exhibit high levels of geographical mobility. As described below, pimp mobility data among the study cities became a key element in the construction of the UCSE proxy.

## 16.1 Previous Methods to Estimate the Size of Underground Economies

Since the early work of Cagan (1958), a number of economists have taken an interest in economic activity that lies outside of measurement. Deemed the "underground economy," this generally includes both legal and illegal activity.[3] Schneider and

[2] We had originally intended also to estimate the size of the UCSE and related domains in Kansas City. However, as we discuss, insufficient data on Kansas City meant that this location had to be dropped from the list of target estimates. For further explanation, see Dank et al. (2014).

[3] See Tanzi (1999) for a discussion of the definitional ambiguities in the "underground economy" literature.

Enste (2000) and Schneider (2011) seek to "provide the reader with some of the latest developments in the size and development of the shadow economies, the driving factors of the shadow economy and the interaction of the shadow economy, with tax morale, government institutions and corruption." To capture underground activity, researchers have developed a number of methods, each with advantages and disadvantages. Interested readers should also see the thorough review in Schneider and Enste (2000) or the more critical review of Tanzi (1999).

Direct approaches use data specifically intended to measure underground activity, such as surveys, interviews, or the results of tax audits. Such approaches are likely to significantly understate the extent of underground activity, since survey respondents may not truthfully report their activity, and audits are limited only to activity that auditors are able to detect. Schneider and Enste (2000) report the results of a large number of methods applied to five OECD countries over a number of decades and find that direct approaches consistently provide the smallest estimates of the size of the underground economy, often only a tenth of other methods' estimates.

In contrast, indirect approaches use a variety of indicators or "proxies," combined with assumptions about their relationships to unobserved activity. For example, building on Cagan (1958), Tanzi (1983) considers currency demand as a proxy. He assumes that (1) all underground activity is conducted using cash, (2) the ratio of currency to official economic activity is constant over time, (3) there exists some base year during which there was no UE,[4] and (4) the velocity of currency is the same in the measured and UEs.[5] From assumption 3, one can calculate the ratio of currency to GDP in the base year. By assumption 2, this ratio does not change over time, so for any subsequent year, we can multiply this ratio by official GDP during that year to calculate how much currency *should* be in circulation due to the legal economy. Any "excess" currency must be devoted to the UE, since assumption 2 ensures that the ratio of currency to official economic activity is constant over time. Assumption 4 then allows us to multiply the currency involved in the UE by the known velocity of money (assumed to be the same between the official economy and the UE) to calculate total underground economic activity associated with physical cash.

Variations on the indirect approach have emerged over time. Other proxy indicators, such as electricity consumption (Kaufman and Kaliberda 1996) and recorded transactions (Feige 1979) have been considered. Like currency demand, these approaches are criticized on the grounds that the relationship between the proxy and the economic activity varies between the legal and illicit sectors and changes over time. In addition, prior proxy-based indirect approaches require identifying a base year (when the UE is presumed not to exist), a process that is fraught with challenges.

[4] The 1930s are often used because of the low income-tax rates.

[5] Our approach uses assumptions 1 and 4, while dispensing with assumptions 2 and 3. Both assumptions 1 and 4 have been criticized in the past. We will discuss our evidence in support of assumption 1 and our attempt to relax assumption 4.

## 16.2 The Estimation Procedure

A naive approach to computing the size of the UCSE might consider the product of (1) estimates of spending by clients and income of pimps and (2) estimates of client and pimp population size. While quantity (1) may be approached through survey interviews, quantity (2) remains elusive and difficult to determine. More significantly, a direct method based on such a product of estimates is sensitive to (the product of) errors in each. By relying instead on a matrix of constraints, the method developed here circumvents such obvious quadratic error sensitivities.

In our initial efforts to arrive at absolute dollar estimates for the size of the UCSE, we began from a tautological observation that we refer to as the law of cash conservation (LoCC):

$$Z(c,t) = S(c,t) + O(c,t)$$

Here $Z(c,t)$, $S(c,t)$, and $O(c,t)$ are, respectively, the total cash in circulation, $Z$, the size of the UCSE, $S$, and the total value of all other cash transactions, $O$, in a fixed city, $c$, in a given year, $t$. If we are able to obtain absolute estimates $Z$ of the left-hand side for two cities $c_1$ and $c_2$, we arrive at two absolute constraints:

$$Z(c_1) = S(c_1) + O(c_1)$$
$$Z(c_2) = S(c_2) + O(c_2)$$

This system of equations, however, contains four unknowns $S(c_1), O(c_1), S(c_2), O(c_2)$. Obviously, a system of two equations and four unknowns is underdetermined, a situation that is only magnified by the addition of other cities. Additional constraints, however, could be derived from the relative size of $S$ and $O$, provided some determinate value could be obtained for their ratios, that is, $S(c_1)/S(c_2)$ and $O(c_1)/O(c_2)$. For example, if we can say that $O(c_2)$ is twice the size of $O(c_1)$ even while $S(c_2)$ is only 1.5 times the size of $S(c_1)$, then we can include these ratio constraints along with the two equations above and arrive at a concrete solution for all variables.

Unfortunately, actual values for $S$ and $O$ are difficult to obtain. Our solution is instead to find proxy measures $S^*$ and $O^*$ that can be assumed to vary linearly with $S$ and $O$, respectively—that is, $S = \theta_s S^*$ and $O = \theta_O O^*$. We note that as the proxies $S^*$ and $O^*$ are only used to determine the numerical ratios in relative constraints, we don't need to know the precise value of $\theta_s$ or $\theta_O$. It suffices that we find a linear relationship between the measurable proxy (e.g., $S^*$) and the unmeasurable quantity it stands for (e.g., $S$) to be plausible and that we are willing to hypothesize that this relationship remains uniform across cities and across time.

While conceptually sound, this approach faces several impediments. First, the requirement of choosing two cities in order to arrive at absolute estimates of $S$ for both cities simultaneously creates a range of combinatorial options, potentially producing discrepant estimates depending on the city pairings considered and the fidelity of their proxy measures $O^*$ and $S^*$. Beyond this, the small size of the UCSE in comparison to the size of non-UCSE makes it sensitive to small perturbations in the ratios of $S^*$ (which are prone to error). This can lead to a high variability in the value of $S$ obtained by solving the simultaneous system of constraints. To address these concerns, several additional steps were added to the estimation.

First, we sought to reduce the variability of estimates for $S$. Toward this, we began by adding variables for two other elements of the underground economy that rely heavily on cash: the trade in illegal guns $W$ and the trade in illegal drugs $D$. The LoCC was thus updated to read:

$$Z(c,t) = S(c,t) + W(c,t) + D(c,t) + O(c,t)$$

The quantities $S$ and $O$ are difficult to obtain in an absolute sense and so it is helpful to constrain them through ratios of aggregate proxies $W^*$ and $D^*$. Assuming we could obtain absolute estimates $Z$ of the left-hand side, each invocation of the LoCC is an absolute *equality* constraint on the sum of four variables.

Second, we added slack to our relative constraints, allowing the proxy measures to exhibit bounded deviations from uniform linear proportionality to the underlying variables. In particular, the $S^*$ and $O^*$ proxies were allowed to deviate from being linearly proportional to $S$ and $O$ by up to $\epsilon_S = \epsilon_O = 20\%$, yielding slack-related four *inequalities*:

$$S(c_2,t)\frac{S^*(c_1,t)}{S^*(c_2,t)}(1/1.20) \leq S(c_1,t) \leq S(c_2,t)\frac{S^*(c_1,t)}{S^*(c_2,t)}(1.20)$$
$$O(c_2,t)\frac{O^*(c_1,t)}{O^*(c_2,t)}(1/1.20) \leq O(c_1,t) \leq O(c_2,t)\frac{O^*(c_1,t)}{O^*(c_2,t)}(1.20)$$

In the case of $S^*$, this 20 percent slack allowed for the possibility of reporting errors in the data collected from UCSE participants. Although we were more certain in our selection of proxies for $D^*$ and $W^*$, we nevertheless incorporated an $\epsilon_D = \epsilon_W = 1\%$ slack for them, yielding four additional slack-related inequalities:

$$D(c_2,t)\frac{D^*(c_1,t)}{D^*(c_2,t)}(1/1.01) \leq D(c_1,t) \leq D(c_2,t)\frac{D^*(c_1,t)}{D^*(c_2,t)}(1.01)$$
$$W(c_2,t)\frac{W^*(c_1,t)}{W^*(c_2,t)}(1/1.01) \leq W(c_1,t) \leq W(c_2,t)\frac{W^*(c_1,t)}{W^*(c_2,t)}(1.01)$$

Third, we sought to circumvent the requirement of choosing two cities at a time for the analysis, by opting to simultaneously consider all possible pairs chosen from seven cities at two times: 2003 and 2007. Since each pair gives us eight linear slack-related inequality constraints, collectively we obtain $\binom{14}{2}8 = 91 \cdot 8 = 728$ slack-related inequalities constraining $14 \cdot 4 = 56$ variables. The solution set of these 728 inequalities was then searched for the point that (a) minimizes the total deviation from the $7 \cdot 2 = 14$ absolute equality constraints arising from the LoCC for each city and time and (b) assigns positive values to all 56 variables. The existence and uniqueness of such a point follow from the theory of linear programming (Dantzig 1947).

## 16.3 The Creation of UCSE Proxy

To construct the UCSE proxy, the study conducted interviews with UCSE participants in eight study locations. The number of interviews in each city is given in table 16.1 for each participant category.

Qualitative data drawn from the subject interviews revealed that child pornography makes a very minor contribution to the size of the sex economy in the United States (see Dank et al., 2014). Thus, in constructing the proxy measure for the UCSE, we focused on scaling up individual weekly pimp revenue, as reported by pimps and sex workers, and disregarded child pornographers.[6]

Although the number of pimp/trafficker interviews was relatively small (N = 73), in the analysis we have chosen to split the data at the median case year (2005), because our aim is to demonstrate how the new methodology presented here is capable of yielding information about the absolute changes in the absolute sizes of UEs over time. A simpler yet analogous procedure that does not split the data set is certainly supported but produces merely a snapshot of economic activity. In making the choice to split the cases for the analysis in this exposition, we have had to reduce the statistical power of our data, and as a consequence, the reliability of our estimates and conclusions is lower than if we had sought merely to demonstrate economic activity at a single point in time.

The interviews also resulted in significant quantitative data, of which the most relevant to this analysis are estimates of the weekly cash income of pimps (see table 16.2). The UCSE participant income reports collected were split at year 2005, with median activity in the bifurcate results located at years 2003 and 2007, thus enabling us to make inferences about longitudinal changes in the cash intake per pimp over the interval in question (2003 to 2007).

**Table 16.1 Subject interviews by participant category.**

| | Pimps/Traffickers | Child Pornographers | Sex Workers |
|---|---|---|---|
| Denver | 7 | 5 | 11 |
| Washington, DC | 6 | 3 | 8 |
| San Diego | 17 | 3 | 0 |
| Miami | 7 | 3 | 0 |
| Seattle | 3 | 3 | 3 |
| Dallas | 15 | 7 | 6 |
| Kansas City | 4 | 6 | 0 |
| Atlanta | 13 | 2 | 8 |
| Oversample | 1 | 1 | 0 |
| Total | 73 | 33 | 36 |

[6] The process of determining weekly pimp incomes is detailed in Dank et al. 2014.

**Table 16.2 Mean weekly cash intake per pimp.**

| City | Pre-2005 Cases | Pre-2005 mean | Post-2005 Cases | Post-2005 Mean |
|---|---|---|---|---|
| Atlanta | 4 | $25,875 | 9 | $32,833 |
| Dallas | 4 | $11,625 | 6 | $12,025 |
| Denver | 3 | $39,333 | 5 | $31,200 |
| Kansas City | 1 | $50,000 | 2 | $5,000 |
| Miami | 7 | $21,929 | 11 | $17,741 |
| San Diego | 4 | $16,500 | 7 | $11,129 |
| Seattle | 2 | $6,750 | 4 | $18,000 |
| Washington DC | 5 | $16,700 | 4 | $11,588 |

In what follows, we describe how this observation was leveraged to design a proxy for the size of the pimp population $G$. This was necessary in order to scale up the mean per-pimp income at each location (table 16.2) to the city scale, providing a proxy $S^*$ for the size of the UCSE at each location. As noted previously, our sample consisted entirely of arrestees, which significantly complicated the estimation of the size of the local pimp population. Were we to assume that the size of the arrestee population was proportional to the overall size of the local pimp population, we would likely overcount those locations with a strong law-enforcement stance toward sex-economy participants. This, in turn, would potentially produce a city- or time-specific law-enforcement-related bias on the values of $G$, violating the defining requirement of a proxy measure: that it vary in uniform proportion to the variable it claims to represent.

We noticed that both pimps and sex workers exhibited high degrees of mobility, often participating in the UCSE across a range of locations in the course of their involvement in it. In what follows, we describe how this mobility was leveraged toward the design of a proxy for the size of the pimp population that is free of systematic law-enforcement-related bias. Although our interview subjects were either incarcerated or part of a court diversion program at the time of the interview, our respondents reported a diverse range of experiences engaging in the commercial sex market. Thus, our data were not biased toward individuals who worked in only one sector of the market (e.g., the streets) but included sex workers and pimps who worked in the streets, online, at massage parlors, in brothels, and in clubs. Often, the respondents worked in multiple venues in order to maximize their profits and clientele.

To avoid the differential effects of variation in local law-enforcement intensity, we defined a variable called "home" $h(p_i) \in C$ corresponding to the city named by a participant as his or her primary place of engagement with the sex economy and based our estimate of the pimp population of any city on the number of pimps who worked in a city but did *not* call that city home, $L(p) \subset C$. From this vantage point, arrests in a city are seen as the outcome of a random capture process that apprehends pimps and sex workers whose mobility corresponds to differing levels of attractiveness to those cities other than their "home" city (which was quite often the city in which they were arrested). Using data from pimps who worked in but did not call a city "home," $h^{-1}(c_j) \in C$ also enlarges the overall sample pool used to determine the size of the UCSE for each city, which in turn serves to mitigate the effects of differential arrest rates through aggregation across broader geographical locations. Given this, a proxy for the size of the pimp population in city $c_i$ can then be taken to be

$$G_1^*(c_j) \stackrel{\text{def}}{=} \frac{|V(c_j)|}{\sum_{c \in N(c_j)} |L(c)|}$$

Here $P = \{p_1, p_2, \ldots, p_N\}$ are the pimps and sex workers who participated in the interviews ($N = 109$), $C = \{c_1, c_2, \ldots, c_n\}$ is the set of cities ($n = 7$), and $L(p) \subset C$ is the set of all study cities excluding a participant's "home" city. $V(c_j)$ is thus the set of those pimps and sex workers who worked in a study city without identifying it as his or her home city during the time period covered by the interview:

$$V(c_j) \stackrel{\text{def}}{=} \{p \in P \mid c_j \in L(p)\}$$

And finally, $N(c_j) \stackrel{\text{def}}{=} P \backslash h^{-1}(c_j)$ are the UCSE study participants for whom city $c_j$ is not home. Defined in this manner, $G_1^*(c_j)$ measures the proportion of UCSE study participants who "visited" $c_i$ for work as a fraction of the total number of "visits" made by UCSE participants whose home was not city $c_j$, because $G_1^*(c_j)$ disregards pimps and sex workers who were captured in city $c_j$ on both the numerator and denominator of its definition and is thus free of city- (or time-) specific law-enforcement-related bias.

Because the $G_1^*$ proxy leverages mobility patterns, it is potentially susceptible to differential effects of variation in the geographical distances between cities. This might lead to a concern regarding the potential for a distance-related bias in the values of $G_1^*$. To address this, we envision a stochastic process that governs the movement of each pimp over time, which collectively determines a stationary distribution of the numbers of pimps in our sample across cities. Within this stochastic process, each city $c_j$ is taken to have some exogenously determined characteristic $M(c_j)$ that serves to attract pimps, enticing each UCSE participant to visit, despite the inconveniences implicit in the distance separating the pimp's home and $c_i$. The value of $M(c_j)$, once determined, would serve as a proxy measure for the size of the pimp population in city $c_j$. At this stage in the analysis, we discovered that very few pimps/traffickers in our sample had ever participated in the UCSE in Kansas City, such that no reliable estimate of the size of the Kansas City UCSE was possible. As a result, Kansas City was dropped from the estimation process; the analysis that follows proceeds with just seven cities.

To determine the value of $M(c_j)$, let $Y_{ij}$ be a 0/1 indicator variable encoding whether UCSE pimp participant $p_i$ decided to work in city $c_j \in C'$. We express $Y_{ij}$ via a logistic model

$$\Pr(Y_{ij} = 1) = p_{ij} = \frac{e^{A_{ij}}}{1 + e^{A_{ij}}}$$

where $A_{ij}$ is an "attraction force" between UCSE participant $p_i$ and city $c_j$ expressed by the proportionality

$$A_{ij} \propto \frac{M(p_i) \cdot M(c_j)}{d(h(p_i), c_j)^{\lambda}}$$

which follows the formal structure of the gravitational law in Newtonian mechanics. Here the "mass" of a UCSE participant $i$ is $M(p_i)$, which is his or her propensity to work in multiple cities, estimated as

$$M(p_i) \approx \left|L(p_i)\right|$$

and the "mass" of a city $M(c_j)$ is its propensity to draw UCSE participants toward it; $d(h(p_i), c_j)$ is the highway distance between the home city of a UCSE participant $p_i$ and the city $c_j$, and $\lambda$ is the decay exponent of the attraction force (as a function of distance $d$).

We estimated the values of $M(c_j)$ and $\lambda$ by a regression analysis of the survey data. Toward this, for each UCSE participant $p_i \in P$ and each of their six nonhome cities $c_j \in C \setminus \{h(p_i)\}$, we introduce a nine-tuple $\mathbb{X}_{ij} \in \mathbb{R}^9$

$$\mathbb{X}_{ij} = \left(X_1, X_2, \ldots, X_7, \ -\ln d(h(p_i), c_j), \ \ln\left|L(p_i)\right|\right)$$

The first $n = 7$ coordinates $X_1, X_2, \ldots, X_7$ are all 0 with the exception $X_j = 1$; coordinate 8 is the natural logarithm of the distance between $h(p_i)$; and $c_j$ coordinate 9 is the natural logarithm of our estimate for $M(p_i)$, which can be thought of as a sex worker fixed effect capturing different individuals' propensity to travel. The resulting $N(n-1) = 109 \cdot 6 = 763$ points are each augmented by $Y_{ij}$ taken as coordinate 10, producing the 10-dimensional dataset

$$\mathcal{D} = \left\{(\mathbb{X}_{ij}, \mathrm{Y}_{ij}) \mid p_i \in P \text{ and } c_j \in C \setminus h(p_i)\right\}$$

This set is then subjected to standard logistic regression to model the probability that coordinate 10 (the dependent variable) $Y_{ij} = 1$ has a linear function of the first 9 (independent variable) coordinates. The coefficient obtained for coordinate 8 is the value $\bar{\lambda} = 1.5$, which represents the estimated distance-decay exponent of the attraction force. We denote the coefficients obtained for the first $n = 7$ coordinates $X_1, X_2, \ldots, X_7$ via this regression as $\theta_1, \theta_2, \ldots, \theta_7$. Inverting the initial logistic transformation of variables needed to carry out the logistic regression, these coefficients, in turn, give rise to the "mass" values $M(c_1), M(c_2), \ldots, M(c_7)$ or $M(c_j)$ of corresponding cities $c_1, c_2, \ldots, c_7$ as listed in the middle column of table 16.3.

Finally, normalizing the $M(c_j)$ values by the median entry (specifically, Washington, DC), we obtain $G_1^*(c_j)$ as a proxy measure for pimp population size. The values of $G_1^*(c_j)$ are listed in the last column of table 16.3.

The results shown in table 16.3 are a significant step toward estimating the relative UCSE participation levels in each city. From this table, we may, for example, learn that the pimp population participating in the UCSE in Atlanta is 19.2 percent larger than

Table 16.3 Pimp population proxy (left); normalized by DC (right).

| City | $M(c_j)$ | $G_1^*(c_j)$ |
|---|---|---|
| Atlanta | 6.254 | 1.192 |
| Dallas | 6.079 | 1.108 |
| Denver | 3.379 | 0.156 |
| Miami | 8.049 | 1.785 |
| San Diego | 5.811 | 0.975 |
| Seattle | 5.463 | 0.805 |
| Washington, DC | 5.861 | 1.000 |

Table 16.4 UCSE proxy $S^* = G_1^*(c_j)$ times mean weekly cash intake per pimp.

| Year | Atlanta | Dallas | Denver | Miami | San Diego | Seattle | Washington, DC |
|---|---|---|---|---|---|---|---|
| 2003 | 30,835 | 12,877 | 6,117 | 39,141 | 16.090 | 5,433 | 16,700 |
| 2007 | 39,128 | 13,320 | 4,852 | 31,666 | 10,852 | 14,489 | 11,588 |

the corresponding population in Washington, DC. Multiplying the proxy measure for pimp population size $G_1^*(c_j)$ with corresponding absolute estimates of the mean weekly gross cash intake by pimps (see table 16.2), we obtain a proxy measure for the size of the UCSE in each year. These product values, listed in table 16.4, scale the relative estimates of pimp-population size by the city-specific amount of cash associated with UCSE participation, which, based on participant reports, varies widely across cities and times.

From table 16.4 we may, for example, learn that the size of the UCSE in Atlanta grew 27 percent between 2003 and 2007. While table 16.3 showed that pimp participation in Atlanta was approximately 20 percent larger than in Washington, DC, table 16.4 shows that the relative size of the UCSE economy in Atlanta in 2003 was $30,835/16,700 = 1.85$ times the size of Washington, DC, in that same year. By 2007, the ratio had grown to 3.38.

## 16.4 The Total Cash Economy and Remaining Proxies

So far, we have determined the *relative* sizes of the UCSE across each of the seven cities in 2003 and 2007. To determine the *absolute* size of each city's UCSE, however, we must begin by determining the size of the overall cash economy in each city ($Z$). Because city-level currency data were not available, they had to be estimated using national data. A first strategy is simply to multiply the national currency-to-GDP ratio by city-level GDP, concluding that the cash in circulation in city $c$ at time $t$ is

$$Z(c,t) = \frac{Z(National, t)}{GDP(National, t)} GDP(c, t)$$

The cash in circulation is estimated by scaling MSA GDP to MSA cash in circulation, using the St. Louis Federal Reserve Bank's Federal Reserve Economic Data (FRED) on the ratio of national GDP to national cash in circulation, while taking into account economic variables that influence the conversion rate. Such a methodology can necessarily only be as robust as the FRED data itself. To the extent that the figures published by FRED may be based on potentially incorrect assumptions concerning a uniform velocity of money (across legal and illegal economies), the estimates we derive here are also prone to the ramifications of the failure of such assumptions.

Further consideration reveals that the currency-to-GDP ratio varies over time and is likely influenced systematically by economic conditions, such as mean per-capita personal income, employment rate, inflation rate, and so on. To quantify this systematic variation, we used quarterly data from FRED for all available years (1959–2012). We used a simple linear model to estimate systematic covariation between economic variables and the national currency-GDP ratio. Experimenting across a broad range of economic variables, we arrived at the parsimonious model shown in table 16.5.

Applying this fitted model of the national currency-GDP ratio as a predictive model in the context of city-level economic variables, we obtained city-level currency-GDP ratios for each city in 2003 and 2007. By using the model, the city-level cash estimates produced take into account variation in economic conditions associated with currency level, both over time and across cities. In practice, there was not much difference between these regression-based estimates and a simpler method that applies the mean national currency-to-GDP ratio (0.0574 over the years 2000 to 2012). Table 16.6 lists the two estimates; in what follows, we use the regression-based estimate.

**Table 16.5 Modeling the influence of economic factors on the currency-GDP conversion rate.**

| Variable | Coefficient Estimate | Standard Error | $t$-Statistic |
|---|---|---|---|
| Intercept | 0.319*** | 0.052 | 6.18 |
| Per-capita personal income | −6.96*** | 1.507 | 4.62 |
| Per-capita real GDP | 6.458*** | 1.309 | 4.94 |
| Ratio of GDP to personal income | −0.227*** | 0.048 | 4.68 |
| Employment-population ratio | −0.084*** | 0.013 | 6.5 |
| Business income as percent of total income | 0.28*** | 0.03 | 9.48 |
| Inflation rate | −0.032*** | 0.007 | 4.63 |

$N = 215$, $R^2 = .902$, $adj.R^2 = .899$, $*p < .10$, $**p < .05$, $***p < .01$

**Table 16.6 Total cash in circulation by regression model and simple linear scaling.**

| City | Year | Regression Estimate (Billions of 2005 Dollars) | Simpler Estimate (Billions of 2005 Dollars) | Ratio |
|---|---|---|---|---|
| Atlanta | 2003 | 12.57 | 13.21 | 0.95 |
| | 2007 | 13.82 | 14.78 | 0.94 |
| Dallas | 2003 | 19.59 | 16.56 | 1.18 |
| | 2007 | 24.23 | 19.5 | 1.24 |
| Denver | 2003 | 8.91 | 7.25 | 1.23 |
| | 2007 | 9.18 | 9.97 | 1.15 |
| Miami | 2003 | 10.31 | 12.26 | 0.84 |
| | 2007 | 10.85 | 14.28 | 0.76 |
| San Diego | 2003 | 7.79 | 7.97 | 0.98 |
| | 2007 | 8.49 | 8.98 | 0.94 |
| Seattle | 2003 | 10.37 | 10.06 | 1.03 |
| | 2007 | 13.05 | 12.02 | 1.09 |
| Washington, DC | 2003 | 17.83 | 18.12 | 0.98 |
| | 2007 | 20.81 | 20.63 | 1.01 |

For all of the variables in the study (except the UCSE values of $S^*$ discussed previously), multiple "atomic" proxies were aggregated to construct a composite proxy (i.e., $W^*$, $D^*$, $O^*$) that stands for and varies in linear proportion to an underlying, unmeasurable quantity of interest (i.e., $W$, $D$, $O$). For example, a good atomic proxy for the size of the cash economy involved in illegal drugs ($D$) would be any measurable quantity whose size can be expected to go up or down in proportion with the growth and shrinkage of the cash-based drug economy. We note that as all our composite proxies are only used to determine the numerical ratios in relative constraints, we don't need to know the precise relationship of the unmeasurable quantity to the proxy measure's value. It suffices that we find a linear relationship between the measurable proxy and the unmeasurable quantity it stands for to be plausible and that we are willing to hypothesize that this relationship remains uniform across cities and across time.

In constructing the composite proxies $W^*$, $D^*$, and $O^*$, we combined a number of separate measurable quantities (see table 16.7), each of which was large enough to be robust against stochastic noise, and to the extent possible, decoupled from fluctuations in law-enforcement policies that we know to be significant across geography and time.

We selected the set of atomic proxies from a broad range of social spheres, which appeared to be independent of one another in all aspects except for their mutual sensitivity to fluctuations in the size of the unmeasurable quantity itself. In this manner, we sought to make the composite proxies robust against imperfections in the respective atomic proxies that might have otherwise eroded our uniformity hypothesis. For example, in choosing both self-reported illicit drug use by youth ages 12+ and emergency-room visits attributed to drug use, we selected two knowable quantities that

**Table 16.7 Atomic proxies for drugs (D1–D7), weapons (W1–W3), and other cash (O1–O3).**

| Data Set | Variables (for years 2000–2010) |
|---|---|
| Illicit Drugs (D) | |
| NSDUH | D1: Illicit drug use in the past month among population ages 12+ |
| | D2: Illicit drug use (other than marijuana) in the past month among population ages 12+ |
| | D3: Cocaine use in the past year among population ages 12+ |
| YRBSS | D4: Percent of high school students offered, sold, or given drugs at school multiplied by total size of respective metropolitan statistical area (MSA)* |
| DAWN | D5: Emergency-room visits attributed to drug use |
| | D6: Emergency-room visits at which drug use was mentioned |
| ADAM | D7: Percent of property-crime arrestees who tested positive for any drug scaled by MSA population |
| Illegal Guns (W) | |
| YRBSS | W1: Percent of high school students who carried a gun multiplied by total MSA population |
| ATF | W2: Number of weapons seized by ATF |
| NVSS | W3: Fraction of suicides committed with firearms multiplied by MSA population |
| Other (O) | |
| N/A | O1: Percent of employment in private-sector service industry, scaled by MSA population |
| | O2: Percent of employment in construction, scaled by MSA population |
| | O3: Percent of employment in food services and drinking establishments, scaled by MSA population |

were expected to vary uniformly across the study cities with the amount of cash spent in each city in the illegal-drug economy. Further, as we expect the number of children ages 12+ using drugs to be independent of the number of emergency-room visits attributed to drug use in a given city at a given time, the use of these two proxies helped ensure that idiosyncratic sensitivities of either proxy were buffered by the presence of the other.

The estimation of the aggregate proxies $W^*$, $D^*$, and $O^*$ for each city depended on seven atomic drug proxies, three atomic proxies for guns, and three atomic proxies for all other uses of cash. The use of multiple atomic proxies obviously requires aggregation into a single composite proxy. To do this, we again used a simple linear regression, which was carried out for all aggregates.

For concreteness, we consider the construction of the drug proxy $D^*$ from the atomic proxies D1 through D7. Here, let $D_k(c,t)$ be the value of the $k^{th}$ atomic proxy for drug activity, measured in city $c = C_j$ at time $t$(where $j = 1, 2, \ldots, 7$, and $t = 2000, 2001, \ldots, 2010$). We begin by normalizing each atomic proxy's value and standard error, by the city-year having the most complete data (Miami in 2003). Thus, we define

$$\tilde{D}_k(c_j,t) = \frac{D_k(c_j,t)}{D_k(c_3,2003)}$$

$$\tilde{\sigma}_k(c_j,t) = \frac{\sigma_k(c_j,t)}{D_k(c_3,2003)}$$

and then estimate a simple linear model

$$\tilde{D}_k(c_j,t) = \hat{\alpha}_j + \hat{\beta}_j t$$

where each observation is weighted by $\frac{1}{\tilde{\sigma}_k(c_j,t)}$. In this manner, we arrive at a time trend for the drug market in each city-year, which incorporates the time trends of each atomic proxy, normalizing their disparate units, while being robust to missing atomic proxy values (for specific city-years) and desensitized to outliers in atomic proxy values and different atomic proxies.

Based on these regression-model coefficients $\hat{\alpha}_j$ and $\hat{\beta}_j$, we calculate $D^*(c_j,t)$ as $D^*_k(c_j,t) = \hat{\alpha}_j + \hat{\beta}_j t$. The computed values of the drugs proxy $D^*$ are listed in table 16.8.

We followed an identical procedure to aggregate the three atomic proxies $W_1 - W_3$ and $O_1 - O_3$ onto the composite $W^*$ and $O^*$ in the case of $O^*$normalizing by Miami, 2007, instead of Miami, 2003, because it had more complete data. The computed values of the guns proxy $W^*$ are given in table 16.9.

Similarly, the values for $O^*$ are shown in table 16.10.

The final set of constraints necessary for the estimation of various economy sizes is a set of positivity constraints that require all values to be well separated from zero. To ensure this, we added 56 additional lower-bound inequalities satisfying the equation

$$v(c_j,t_i) \gg \gamma\, \tilde{Z}(c_j,t_i)$$

**Table 16.8 Composite proxies for illegal drug $D^*$ economy.**

| Year | Atlanta | Dallas | Denver | Miami | San Diego | Seattle | Washington, DC |
|---|---|---|---|---|---|---|---|
| 2003 | 1.038 | 1.334 | 0.544 | 0.931 | 1.041 | 0.870 | 1.099 |
| 2007 | 1.153 | 1.879 | 0.630 | 0.944 | 0.949 | 0.862 | 1.007 |

**Table 16.9 Composite proxies for illegal gun $W^*$ economy.**

| Year | Atlanta | Dallas | Denver | Miami | San Diego | Seattle | Washington, DC |
|---|---|---|---|---|---|---|---|
| 2003 | 1.293 | 1.307 | 0.446 | 0.807 | 0.356 | 0.628 | 1.135 |
| 2007 | 1.123 | 1.317 | 0.362 | 0.913 | 0.368 | 0.495 | 1.221 |

**Table 16.10 Composite proxies for the economy of all other cash-based expenditures.**

| Year | Atlanta | Dallas | Denver | Miami | San Diego | Seattle | Washington, DC |
|---|---|---|---|---|---|---|---|
| 2003 | 0.961 | 1.121 | 0.496 | 1.019 | 0.564 | 0.643 | 0.979 |
| 2007 | 1.120 | 1.324 | 0.481 | 1.027 | 0.568 | 0.695 | 1.179 |

with $\gamma = 1/1000$ and $v \in \{S, D, W, O\}$ for each of the individual city estimates in 2003 and 2007.

## 16.5 The Estimates

Taken together, the above considerations create a system of 798 constraints on 56 variables (728 slack-related inequalities, 14 absolute constraints from the LoCC, and 56 positivity constraints) that must be simultaneously satisfied. Assuming that the feasible set for this system of constraints is nonempty, there are likely to be infinitely many solutions. This is because the nonempty intersection of 798 half-spaces is, in general, a convex polytope in $\mathbb{R}^{56}$. To specify a unique solution from within the feasibility set, we must specify an objective function $E$ that we wish to minimize over the feasible set. The objective function allows us to find the "best" solution point that satisfies the constraints. We wish to take $E$ to be the average relative divergence from the modified version of the LoCC

$$Z(c,t) = S(c,t) + W(c,t) + D(c,t) + O(c,t)$$

And so we define

$$E \stackrel{\text{def}}{=} \sum_{c_j \in C;\, i \in \{0,1\}} \left| 1 - \frac{S(c_j,t_i) + W(c_j,t_i) + D(c_j,t_i) + O(c_j,t_i)}{Z(c_j,t_i)} \right|$$

We note that the $E$ we seek is given by the point within the feasible solutions that minimizes the normalized discrepancy between the estimate of the size of the total cash economy and the sum of the individual subeconomy estimates. The uniqueness of the solution point that minimizes $E$ follows from the theory of linear programming (Dantzig 1947). The values of the 56 variables at the optimal solution point are listed in table 16.11.

To compute the solution (minimize $E$, subject to 789 linear constraints in $\mathrm{R}^{56}$), we used the noncommercial industry standard linear program solver *lp_solve*, written in ANSI C by Michel Berkelaar, which is known to efficiently solve systems of linear equations with up to 30,000 variables and 50,000 constraints (Berkelaar et al. 2014).

Table 16.11 Estimates of the absolute sizes of subeconomies (millions of 2005 dollars).

| City | Year | Sex | Drugs | Guns | Other |
|---|---|---|---|---|---|
| Atlanta | 2003 | $238 | $104 | $169 | $14,500 |
| | 2007 | $290 | $117 | $146 | $16,000 |
| Dallas | 2003 | $99.4 | $134 | $171 | $16,900 |
| | 2007 | $98.8 | $191 | $171 | $19,000 |
| Denver | 2003 | $47.2 | $54.7 | $58.4 | $7,470 |
| | 2007 | $39.9 | $63.9 | $47.4 | $7,820 |
| Miami | 2003 | $302 | $93.4 | $106 | $15,300 |
| | 2007 | $235 | $95.7 | $118 | $14,700 |
| San Diego | 2003 | $124 | $105 | $46.6 | $8,490 |
| | 2007 | $96.6 | $96.3 | $47.7 | $8,740 |
| Seattle | 2003 | $50.3 | $87.3 | $83.1 | $9,840 |
| | 2007 | $112 | $87.4 | $60.1 | $11,800 |
| Washington, DC | 2003 | $155 | $111 | $150 | $17,700 |
| | 2007 | $103 | $103 | $160 | $20,300 |

## 16.6 Discussion

The 56 entries in table 16.11 can be used to ascertain various facts about time-based trends in underground cash-based economies of the seven cities. For example, the underground cash-based drug economy increased in Atlanta, Denver, and Dallas, while the underground cash-based gun economy decreased in Atlanta, Denver, and Seattle. At the same time, the size of the UCSE is seen to increase significantly in Atlanta and Seattle, hold steady in Dallas and Denver, and fall in Miami, Seattle, and Washington, DC. While we can estimate the absolute magnitudes of these changes, the new methodology is necessarily silent about possible reasons underlying the different ways these metropolitan UEs evolved in the course of the interval from 2003 to 2007.

With regard to the UCSE, it is important to note that the estimates generated are of the total UCSE, not UCSE spending per person. Given that the figures are not adjusted for population, part of the reason Dallas has 2.5 times the activity of Denver in 2007 is that the population of the Dallas metropolitan area was 2.5 times that of the Denver metropolitan area. However, the numbers do not simply reflect population differences. For instance, although Miami and Washington, DC, have approximately the same metropolitan area population (5.5 million and 5.3 million, respectively), by our estimates, Miami has about twice as much UCSE activity as Washington, DC.

Limitations of the present exposition revolve around evaluating the significance of the observed increases or decreases (e.g., is the 1 percent drop reported in the Dallas UCSE between 2003 and 2007 significant, relative to the errors in the estimation procedure inputs?). Answering such concerns precisely requires, first, formally modeling our *assumptions* about the fidelity with which each of the 14 proxies tracks the corresponding unmeasurable quantity. For example, how closely do we believe the D5 proxy (emergency-room visits attributed to drug use) tracks actual cash spent on drugs? Second, with these 14 parameterized formal models of proxy fidelity in hand, we can analyze the sensitivity of the dependence of each of the 56 outputs presented in table 16.11 to the free parameters in the proxy fidelity models. Given the complexity of the linear program at the core of the estimation procedure, such an analysis must necessarily be carried out through stochastic sampling and simulation. Finally, by aggregating the results of stochastic simulations, we arrive at a joint sensitivity analysis, allowing us to make assertions of the form "If we believe that the proxies track the underlying unmeasurable quantities within a factor of $\epsilon_1, \epsilon_2, \ldots, \epsilon_{14}$, then our confidence in the estimates produced in table 16.11 is $\delta_1, \delta_2, \ldots, \delta_{56}$ of their true values."

While a complete exposition of the subtleties underlying the above multifactor sensitivity analysis is beyond the scope of this chapter, we can provide the reader with a sense of the solution quality and the broad reliability of the final estimate figures. Toward this end, we recall that the objective function $E$ at the optimal solution point takes a value $E = 0.0576$, implying that for the values given in table 16.11, on average, each of the 14 constraints mandated by the LoCC experienced a 5.76 percent deviation from perfect equality.[7] In the course of the analysis, certain parameters were chosen, including the slack margins $\epsilon_S, \epsilon_D, \epsilon_W, \epsilon_O$ that were used to relax each strict equality based on proxy ratios into a pair of linear inequalities. The settings of these slack margins are shown in table 16.12.

The slack margins above specify the apex angles of the 728 "wedges" (in $\mathbb{R}^{56}$) corresponding to each of the 728 slack-related *inequalities*, where each wedge is centered along the hyperplane induced by a proxy-ratio *equality* constraint. From a geometric interpretation of these inequalities, one sees that the choice of slack margin assumptions (which govern the apex angles) is closely related to proxy fidelity

**Table 16.12 The amount of slack permitted each of the variables.**

| Slack | Setting |
|---|---|
| $\epsilon_S$ | 0.20 (20%) |
| $\epsilon_D$ | 0.01 (1%) |
| $\epsilon_W$ | 0.01 (1%) |
| $\epsilon_O$ | 0.20 (20%) |

[7] That is, on average, in each of the 14 LoCC equations, the sum of the right-hand side UE variables is <6% away from known absolute total cash $Z$ on the left-hand side.

Table 16.13 The impact of slack margin on the error.

| $\epsilon_O$ and $\epsilon_S$ | E |
|---|---|
| 0.01 (1%) | 18.83% |
| 0.05 (5%) | 15.35% |
| 0.10 (10%) | 11.64% |
| 0.15 (15%) | 8.54% |
| 0.20 (20%) | 5.76% |
| 0.25 (25%) | 3.57% |

assumptions (which govern the central axis of each wedge). Table 16.13 shows how solution quality, measured by the value of the objective function $E$ is affected when $\epsilon_O$ and $\epsilon_S$ are changed.

From table 16.13, we see a clear trade-off between the quality of the solution and the slack margins for the sex and other proxy ratios. Assuming fidelity of 20 percent for the sex and other proxies, we obtain a less than 6 percent average deviation in the law of cash conservation. We feel this is quite acceptable, given the fundamental unavoidable assumption implicit in the law: that every dollar expended within region $c$ during a fixed time interval can be unambiguously characterized as having been related to the acquisition of either illegal commercial sex or drugs or guns or "none of the above."

Another limitation of the present exposition is the need to interpret the final estimates (table 16.11) in the social, economic, and historical context specific to the cities and time periods in question. It is natural to ask, for example, about the extent to which the figures in table 16.11, amortized over each corresponding MSA population, yield per-capita expenditures that are commensurate with current estimates derived by other more conventional analytic techniques? We intend to carry out such a cross-validation of estimates in the future, in collaboration with domain experts in each of the underground economies appearing in this analysis.

In closing, the technique presented in this chapter illustrates the methodological promise of combining absolute constraints (here based on total cash in circulation in various cities), together with large numbers of ratio constraints (here based on proxy variables that are assumed to track various unmeasurable quantities). Given a large system of such constraints, one may leverage optimization theory to determine the unique solution point that both satisfies the proxy-based inequalities and most closely meets the specified absolute constraints. In the context of estimating the absolute sizes of otherwise inaccessible quantities such as the size of an underground metropolitan cash-based economy for commercial sex, this constitutes a new technique and a significant step forward.

## References

Berkelaar, Michel, Jeroen Dirks, Kjell Eikland, and Peter Notebaert. 2014. *Lp_solve Reference Guide*. http://lpsolve.sourceforge.net/5.5/. Accessed May 1.

Cagan, Phillip. 1958. "The Demand for Currency Relative to the Total Money Supply." *Journal of Political Economy* 66, no. 4: 303–328.

Dank, Meredith, Bilal Khan, Mitch Downey, Cybele Kotonias, Deborah Mayer, Colleen Owens, Laura Pacifici, and Lilly Yu. 2014. *Estimating the Size and Structure of the Underground Commercial Sex Economy in Eight Major US Cities*. Final report submitted to the National Institute of Justice.

Dantzig, George B. 1949. "Programming of Interdependent Activities: II Mathematical Model." *Econometrica* 17, no. 3: 200–211.

Feige, Edgar L. 1979. "How Big Is the Irregular Economy?" *Challenge* 22, no. 1: 5–13.

Kaufmann, Daniel, and Aleksander Kaliberda. 1996. "Integrating the Unofficial Economy into the Dynamics of Post-Socialist Economies." *Economic Transition in Russia and the New States of Eurasia* 8: 81.

Newton, Phyllis J., Timothy M. Mulcahy, and Susan E. Martin. 2008. *Finding Victims of Human Trafficking*. Washington, DC: National Institute of Justice.

Schneider, Friedrich, ed. 2011. *Handbook on the Shadow Economy*. Cheltenham, UK: Edward Elgar.

Schneider, Friedrich, and Dominik H. Enste. 2000. "Shadow Economies: Size, Causes, and Consequences." *Journal of Economic Literature* 38, no. 1: 77–114.

Tanzi, Vito. 1983. "The Underground Economy in the United States: Annual Estimates, 1930–80. *Staff Papers—International Monetary Fund* 30, no. 2: 283–305.

Tanzi, Vito. 1999. "Uses and Abuses of Estimates of the Underground Economy." *Economic Journal* 109, no. 456: 338–347.