Published in final edited form as:

J Hous Econ. 2016 March 1; 31: 14–34. doi:10.1016/j.jhe.2015.12.001.

Homelessness prevention in New York City: On average, it works

Sarena Goodman^a, Peter Messeri^b, and Brendan O'Flaherty^{c,*}

^aDivision of Research and Statistics, Federal Reserve Board of Governors, 20th and C St NW, Washington, DC 20551, United States

^bDepartment of Sociomedical Sciences, Mailman School of Public Health, Columbia University, 722 W. 168 St., New York, NY 10032, United States

^cDepartment of Economics, Columbia University, Mail code 3308, 420 W. 118th St., New York, NY 10027, United States

Abstract

This study evaluates the community impact of the first four years of Homebase, a homelessness prevention program in New York City. Family shelter entries decreased on average in the neighborhoods in which Homebase was operating. Homebase effects appear to be heterogeneous, and so different kinds of averages imply different-sized effects. The (geometric) average decrease in shelter entries was about 5% when census tracts are weighted equally, and 11% when community districts (which are much larger) are weighted equally. This study also examines the effect of foreclosures. Foreclosures are associated with more shelter entries in neighborhoods that usually do not send large numbers of families to the shelter system.

Keywords

Homelessness prevention; Shelter entries; Family homelessness; New York City; Quasi-experimental; Theory of prevention evaluation

1. Introduction

Homelessness exacts a high cost on wellbeing and health (Aaronson, 2000; Astone and McLanahan, 1994; Salit et al., 1998). Its toll on children is particularly severe, leading to developmental delays, cognitive impairment, and increased mental health problems (Bassuk et al., 1997; Buckner, 2004; Haveman et al., 1991; Mohanty and Raut, 2009; Shinn and Weisman, 1996; U.S. Department of Housing and Urban Development, 2012a). Johnson and Scutella (2014) find that entry into homelessness causes psychological distress, and this distress is greatest on entry. This result suggests that reductions in homeless population caused by reducing entries might be better than equivalent reductions achieved by shortening spells.

*Corresponding author: sarena.f.goodman@frb.gov (S. Goodman), pam9@columbia.edu (P. Messeri), bo2@columbia.edu (B. O'Flaherty).

Most homeless people are single adults; however, approximately 40% are members of families with children (Shinn et al., 1998). Substance abuse and mental illness may be the driving forces for homelessness among single adults, but not for families. Adults in homeless families are no more likely than poor-but-housed families to experience behavioral health problems (Culhane et al., 2007; Culhane and Metraux, 2008). The factors that distinguish homeless families from their housed counterparts—domestic violence, housing instability, strained social networks, and lack of housing and social welfare assistance—suggest a process of becoming unhoused, precipitated by episodic housing emergencies, which are distinct from the chronic conditions associated with single adult homelessness (Culhane et al., 2007).

The temporary shelter system is the primary U.S. response to housing emergencies. For many families, temporary shelter stays last many months (Culhane et al., 2007; Goodman et al., 2014). Culhane and Metraux (2008) point out that sole reliance on emergency shelter systems is an inequitable and seemingly inefficient way to help families with housing emergencies and call for experimentation with a broad range of community support programs. The research on community-based homelessness prevention programs to date outside New York City has largely consisted of uncontrolled studies that cannot estimate credibly how many participants would have become homeless without these programs, or how long they would have stayed homeless (Apicello, 2010; Apicello et al., 2012; Burt et al., 2005; U.S. Department of Housing and Urban Development, 2012b).

In New York City, there is good reason to explore whether community-based programs can avert family shelter entries. Because of litigation and subsequent settlements, New York City provides single units with private bath and kitchen facilities for most families with housing emergencies (Culhane et al., 2007). Hence, family homelessness in NYC is essentially about entry into and length of stay in its large and costly family shelter system (U.S. Department of Housing and Urban Development, 2012a). The average family stays in the shelter system is over a year at a cost of over \$30,000. Devoting resources to low-cost community-based programs that keep families in their homes, if successful, might avoid the disruptive effects of shelter entry on the lives of family members and produce substantial cost savings to the governments that fund shelters.

To explore whether a community-based program that targets services to families in housing emergencies could reduce shelter entries, the NYC Department of Homeless Services (DHS) started Homebase (HB) in November 2004. As the largest, and among the earliest, community-based homelessness prevention programs in the United States, HB served approximately 11,000 families during the four years we studied the program, and currently serves around 10,000 households a year.

In this paper, we examine the effectiveness of Homebase during its first four years of operation. We produce estimates for the community impact of Homebase, i.e., whether the total number of entries from neighborhoods that Homebase served decreased, regardless of whether those entries were Homebase participants or not. This net figure is the bottom-line question for an agency. Our estimates imply Homebase decreased shelter entries by 5–11%.

Ours is one of two evaluations of this program's goals. Rolston et al. (2013) report on a small controlled experiment conducted in 2010 that followed individual families, comparing 150 families who were offered HB services to 145 families who were not. They found that HB families were less likely to enter shelters than control families. We discuss this report in more detail in Section 9. The presence of both a controlled experiment and a natural quasi-experiment provides an opportunity to highlight the strengths and weaknesses of each method.

The question we ask—how did Homebase affect shelter entries in the communities it served?—is different from the question Rolston et al. asked. Their experimental setting required action on the part of families at risk (i.e., they needed to show up at Homebase centers for services), and thus they identify the effect of treatment on the treated within this group. Such an estimate cannot be used to assess the general equilibrium effects of the program, e.g., whether “musical chairs” effects merely divert resources to program participants and change the names of shelter entrants but not their numbers; whether network effects turn one averted entry into several by setting an example or providing helpful information; whether anticipation of assistance leads families to take greater risk and therefore need Homebase assistance more when those risks turn bad (ex ante moral hazard); whether families who were never actually enrolled gain enough helpful information and advice in a short visit to avert homelessness sans formal Homebase services.

Furthermore, no single experiment should be taken as the last word in deciding policy on community-based prevention programs. Potential variation in type and implementation of interventions combined with unique features of local context will likely produce different results in different times and places. The question of whether and why community-based programs prevent family homelessness will rest on accumulating a large evidence base from many types of intervention studies. Quasi-experimental designs are likely to loom large in answering this question, since administrative data on homelessness and shelter entries are easier to collect than setting up a randomized design and pose fewer ethical problems about denying services to communities hard pressed for stable housing.

However, quasi-experiments, such as ours, come with a cost. The analysis is complicated, as more work is required to model plausible counterfactual conditions; the major advantage of the randomized experiment is the simplicity of the ready-made counterfactual. In other words, one weakness of our study is that communities were not randomly assigned to treatment and control groups, which thus requires assumptions about how treated communities might have fared were they not assigned to treatment. (There are also specification issues that we discuss in Section 5.)

Because, in our setting, the true counterfactual is unknowable, we construct several plausible counterfactual conditions and triangulate an HB effect from multiple estimates. For instance, we use two different geographic partitions of New York City's neighborhoods: community districts (CDs) and census tracts (CTs). Census tracts are more plentiful and exhibit greater variation in intensity of HB services, but community districts have cleaner definitions of when HB was operating and because they are larger than census tracts, allow us to net out at

least some of the potential role that HB might play in increasing or decreasing entries by non-participants.

Besides estimating the effects of HB, we also examine the impact of foreclosures on entries to homeless shelters. These are the first results we are aware of that attempt to link foreclosures to homelessness at a community level. Lazaryan et al. (2014) attempt an analysis at the individual level. Specifically, we consider *lis pendens* (LP) filings, the first step in the lengthy foreclosure process in New York. Many LPs do not result in foreclosures,¹ but in those that do the foreclosure and resale process may typically take 12–18 months from LP.

Our preferred estimates indicate that Homebase reduced shelter entries in the (geometric) average neighborhood in which it was fully operating by 5–11%. Homebase caused smaller reductions in the times and places where it was operating partially.

Foreclosures, on the other hand, are associated with more shelter entries in many neighborhoods. However, we have not tried to show causality. We do not find statistically significant effects of foreclosures in general or in neighborhoods with historically above-average flows to the shelter system. However, the long-run semi-elasticity of shelter entries with respect to foreclosures is about 0.04 in neighborhoods where a smaller than average number of families historically entered shelters. The rise in foreclosures between 2005 and 2008 is associated with over a third of the increase in shelter entries that these neighborhoods experienced in this period.

We begin in Section 2 with a description of Homebase and its history. Section 3 describes the study data, and Section 4 presents summary statistics. Section 5 considers how homelessness prevention operates and explains the specifications we use. Section 6 presents basic results for Homebase and foreclosures, and evidence on geographic spill overs. Section 7 further illuminates how Homebase worked by looking at its effect on different classes of CD-months. Section 8 addresses the problems of estimating the number of entries that were averted. Section 9 compares our results to those of the randomized control trial (RCT). Mathematical results and proofs are found in the Appendix.

2. What does Homebase do?

2.1. The target client population: families in housing crisis

Families who are not homeless enter shelters when some shock occurs that makes their current living conditions untenable. See O'Flaherty (2009, 2012) for a detailed discussion. Overcrowding, eviction and domestic violence are the major reasons why families are determined eligible for shelter. At the beginning of the study period, overcrowding was the predominant reason, but its share fell: “This could mean, for example that the family's prior residence lacked adequate rooms for children of different genders to sleep, because there were not enough beds (or couches) for unrelated adults, or the presence of extra furniture to accommodate the doubled-up family presented a fire hazard. In 2002, 2,431 entries into

¹Of *lis pendens* filings in New York City in 2007, only 14% had ended with bank ownership or third party auction by 2009; 54% had had no subsequent legal transactions. See Furman Center 2010.

shelter (38%) were because of overcrowded prior living conditions” (NYC Independent Budget Office, 2014). The vast majority of overcrowded families at this time had been doubling up with the mother's parents. While overcrowding as a reason for shelter entry fell, eviction rose (from 17% of entries in 2002 to 32% in 2010), and became the most common reason in 2007. The number of families eligible because of domestic violence also rose, and domestic violence was the most common reason for eligibility in 2008 (NYC Independent Budget Office, 2014). Thus many families enter shelters because they have problems that some form of short-term assistance might conceivably alleviate.

2.2. Homebase program components

Homebase is designed to help families overcome these immediate problems and obstacles that may cause them to lose housing; and therefore, to reduce family entries into homeless shelters. New York City contracts 13 not-for-profit agencies to operate Homebase Centers. These agencies are experienced providers of case management and comprehensive social welfare services for indigent families. Homebase families experiencing housing difficulties voluntarily apply to the Homebase Center closest to their neighborhood. Homebase case managers have wide discretion in matching services to the specific problems of eligible families. Services include family and landlord mediation, legal assistance, short-term financial assistance, mental health and substance abuse services, child care, and job search assistance. During the study period the typical HB case was opened for about six months, and very few cases were repeat families.

2.3. Homebase program implementation

DHS began Homebase as a pilot project in six of the 59 NYC community districts (CDs). The average CD had a 2010 population of 138,503 or what would be considered a mid-sized city in most of the US. The initial HB-eligible CDs were not chosen randomly. They were selected from among the 14 CDs with the largest number of shelter entries during the immediate pre-program period. We refer to these CDs as the “Big Six.” Fig. 1 shows that the location, key demographic and housing characteristics of the Big Six. They were dispersed over four of New York City's five boroughs. Five of the Big Six were located in CDs with the highest concentration of poverty.

DHS expanded HB citywide in two phases. In July 2007, 31 more CDs were included, and the remaining 22 CDs started in January 2008. The expansions were not random either. The July 2007 expansion was mainly in Manhattan and Queens, and by its end every CD in those two boroughs was operating. The January 2008 expansion was mainly in the Bronx and Brooklyn, but also included all three CDs in Staten Island. Although HB centers before 2008 were supposed to serve only families living in eligible CDs, they did not always abide by this rule. About 6% of HB participants received services before HB was operating in the CDs in which they appeared to live. Our models include a parameter to isolate the HB effect for families residing in ineligible CDs from the effect on officially eligible clients.

During its initial phase through June 2007 there was an emphasis on generating clients through targeted outreach. Outreach began with maps prepared by the Department of Homelessness Services that displayed neighborhoods with concentrations of unstable

housing. Staff then blanketed these communities through informational presentations communities at numerous community meetings. The pool of potential clients was augmented through word of mouth. Toward the end of the study period client recruitment shifted from the community at risk to families. Homebase stationed staff at the City's central shelter intake center—about half of new clients were recruited from the intake center in an effort to divert those families who might best benefit from the Homebase program. The data for this study restricted to families recruited from community and not those diverted from the intake center (Scott Auwater, personal communication).

HB centers are instructed to provide services only to eligible families. Eligibility includes an income criterion (200% of the federal poverty level during our study period) and, through December 2007, a residence criterion. Although HB centers before 2008 were supposed to serve only families living in eligible CDs, they did not always abide by this rule. About 6% of HB participants received services before HB was operating in the CDs in which they appeared to live. Our models include a parameter to isolate the HB effect for families residing in ineligible CDs from the effect on officially eligible clients.

HB staff use their discretion to decide whether applicants experience “treatable emergencies”—that is, whether they would benefit from HB or are better served by referral to other City programs. Ineligible families sometimes received advice and referrals. Families with domestic violence, mental health or substance abuse problems that may otherwise have met HB eligibility criteria were often referred to specialized housing services (Shinn et al., 2013).

During the first three years of operation Homebase served 8294 clients. The typical HB family was a single mother of color engaged in low paid job with variable income (e.g. home health care aide). Homebase's primary service objective is to keep families in current housing. Fewer cases are about changing apartments or doubled up situations. Paying off arrears roughly amounts to \$3000–\$4000 per family (S. Auwater, personal communication). During the four-year study period just under half of HB clients (4248) received “housing location and supplies, rental assistance,” but over a third received “job search, training and education” (3586) and “legal action and entitlements advocacy” (2832). Smaller numbers received “mediation and independent living” (1632) and “family health and child care” (941). Some clients received multiple services (NYC Independent Budget Office, 2008).

For families who visit HB, the relevant alternative to shelter is conventional housing, not street homelessness. Since the number of families living on the street was minuscule throughout the period we study, we believe that Homebase did not increase street homelessness among families.

3. Data

We estimated HB effects on family shelter entries through a retrospective analysis of administrative data provided by DHS that included anonymous listings of family shelter entries between 2003 and 2008, and family applications for HB services between November 2004 and the end of 2008. We do not link HB clients to shelter entrants because our analysis estimates the HB effect on shelter entries for communities, not for individuals. Because of

the way the program interacts with homelessness outcomes, “community” is not a uniquely defined concept, and so we undertook parallel analyses in which CDs and CTs were operationalized as alternative geographic units of analysis.

Our primary dependent variables were monthly counts of family entries into the New York City family homeless shelter system, aggregated by 59 CDs and 1889 CTs. Consistent with DHS definitions, “families,” were any group of individuals who live together, and pregnant women. We include both families with children and families with no children (“adult families”). A measure of HB services was constructed from counts of the number of family HB cases opened each month by the CDs and CTs in which they resided. The measure of HB coverage comes from publicly available information on the dates when each CD became officially eligible. The Furman Center at New York University provided us with *lis pendens* (LP) filings in New York City by date and address for the period 2001–2008.

We stratified CDs and CTs by number of entries during the 22-month pre-program period. Homelessness in New York as in most other cities is concentrated in a relatively small number of impoverished communities (see Fig. 1). The substantial variation in community risk of homeless and the factors that drive this risk lead us to investigate how Homebase effects may vary by level of use of these services. We discuss below in more detail why stratification will help us estimate the effect of Homebase. We assumed that the ordering of neighborhoods by characteristics that determine risk of homelessness—the supply of affordable rental apartments, socioeconomic status of residents, family and employment stability—was relatively stable during the six-year period of this study. Comparing shelter entries from each CT before and after the start of the HB programs supports this assumption. The correlation between the total count of families entering the shelter system by census tract for the 22 months just prior to the start of HB services and the total count for the first four years of HB operations was 0.93.

We divided the CDs into a high-shelter-use stratum that included 21 CDs that experienced an above average number of entries during the pre-program period, and a low-use stratum for the remaining 38 CDs. The CTs were divided into three strata. CTs with 0, 1, or 2 shelter entries during the pre-program period formed a low use stratum, those with 3–18 shelter entries formed a moderate-use stratum, and those with 19–67 shelter entries formed a high-use stratum. Most census tracts, 995, fell into the low shelter use group. There were 702 moderate-use and 192 high-use census tracts. The moderate and high use census tracts were heavily concentrated in high-use CDs, but the correlation was not perfect.

4. Summary statistics

From 2003 to 2008, 45,088 families entered the NYC shelter system, an average of 635 families each month. Fig. 2 shows that the monthly average obscures pronounced seasonal variation coupled with a sharp secular rise in shelter entries. In any calendar year, entries tend to peak in the late summer months and ebb during the winter and early spring. The amplitude of seasonal variation is large, on the order of a 200–300 change in monthly entries. HB began in November 2004 midway through a period of relatively low shelter use that extended through 2005: monthly family entries fluctuated between 400 and 600. After

2005 family entries began to climb. The numbers of monthly entries increased to between 600 and 800 during 2006, 2007, and the first half of 2008. By the second half of 2008, monthly entries jumped to over 1000 as the Great Recession began.

Table 1 shows annual trends in HB operations and trends in number of monthly cases opened. During the first 49 months of HB operations (November 2004–November 2008), HB centers served 10,978 families. During the periods of official operations in a CD, HB opened an average of 9.7 family cases a month compared to 10.7 families entering the shelter system. However the overall averages obscure a substantial dilution in treatment exposure after the program expanded in July 2007. During the first two full years of restricted HB operations, the number of HB cases opened in the Big Six CDs substantially exceeded the number of families entering the shelter system from these CDs. For every 10 shelter entries, approximately 16 HB cases were opened. After June 2007, many more families entered shelters from eligible CDs than received HB services, both because of the expanded geographic coverage of HB services and the rising number of shelter entries. For every 10 shelter entries, only 3.3 HB cases were opened.

5. Theory and estimating strategy

Our empirical approach to studying Homebase and our understanding of the difference between RCTs and natural quasi-experiments are both motivated by theoretical considerations about how homelessness prevention programs work. We are interested in how many entries HB averted, but we cannot estimate that number in isolation: we have to decide whether to estimate entries averted per operating CD-month, entries averted per family served, or entries averted per shelter entry. In this section, we will discuss generally how we arrived at the specifications that we use. The Appendix contains more formal derivations.

5.1. General considerations

We are concerned about ratios between three numbers: entries averted, families served by HB, and shelter entries. The last two are observable, and, as we have seen, the ratio between them varies tremendously during the CD-months of HB operations that we observe. Hence it is impossible for both the ratio of entries averted to families served and the ratio of entries averted to shelter entries to be constant or even moderately stable.

Ignoring for the moment effects on non-participants we can think about four classes of families in any CD-month:

- Families who are not served by HB and who enter the shelter system. We call these families “non-HB entries.”
- Families whom HB serves, but who still enter the shelter system (either because their problems were too severe for HB to resolve, or because HB did not perform adequately). We call these families “treatment failures.”
- Families whom HB serves, who would not have entered the shelter system in the absence of HB, and who do not enter after HB service. We call these “un-needed treatments.”

- Families whom HB serves who would have entered the shelter system but do not because of HB treatment. We call these “averted entries.”

Shelter entries are the sum of treatment failures and non-HB entries. Families served are the sum of treatment failures, un-needed treatments, and entries averted.

Think first about the ratio of *entries averted to families served*. Consider a shock like a recession that increases housing distress in New York City's population. If DHS maintains the same eligibility standard and outreach strategy in the face of the shock, all three components of HB families served will change, with treatment failures likely rising and un-needed treatments likely falling. The ratio of entries averted to the other two components is unlikely to stay constant. (The Appendix gives an example—displaced exponential distribution of housing distress—where the ratio is constant, but that example implies a constant ratio between families served and shelter entries, which the data do not support.)

Similarly consider policy changes by DHS. Raising eligibility standards decreases the number of un-needed treatments and so raises the ratio of entries averted to families served. Increasing outreach activities preserves this ratio only if the new families attracted to HB are divided among the three components the same way as original families were (no selection effect, in other words), which is unlikely. Thus we do not expect the ratio of entries averted to families served to be stable.

The same sort of reasoning applies to the ratio of *entries averted to shelter entries*. A recession may increase entries averted, as well as the two components of shelter entries, non-HB entries and treatment failures (possibly because larger numbers of families coming to HB may reduce effort that staff can devote to each family). However, there is no reason to think that the ratio will not change. Increasing outreach activities, if successful, raises the ratio by reducing non-HB entries and raising entries averted (unless a flood un-needed treatment families reduces the effort staff can devote to other families and increases treatment failures at the expense of averted entries). Changing the eligibility standard will not affect this ratio, but it is only one possible policy option. Once again, we do not expect this ratio to be stable.

The instability of each of these two ratios—entries averted to families served and entries averted to shelter entries—has different implications for our study. Our response to the first ratio—entries averted to families served—is simple: we do not estimate it, even though it is the most commonly discussed ratio (and the ratio that RCTs by necessity must estimate). But we cannot ignore the second ratio—entries averted to shelter entries—because our estimating strategy, difference-in-differences, forces us to consider it. This is because we expect shocks to have a multiplicative effect on shelter entries, not an additive effect, as we describe below.

5.2. Functional form

Intuitively, the substantial seasonal variations in shelter entries and large secular increase due to the onset of the Great Recession should raise entries (in the absence of HB) by more in neighborhoods that previously had many entries than in those that previously had few.

The expectation that neighborhood responses to external shocks are proportional to existing levels of shelter entries and not additive suggests that we should estimate equations that are of a loglinear form.

$$\ln s_{ct} = \beta H_{ct} + \gamma_c + \delta_t + \varepsilon_{ct}. \quad (1)$$

rather than an additive form

$$s_{ct} = \beta H_{ct} + \gamma_c + \varepsilon_{ct}. \quad (2)$$

where s_{ct} denotes shelter entries in neighborhood c and month t , γ_c is a fixed effect for neighborhood c , δ_t is a fixed effect for month t , H_{ct} is a dummy equal to one if and only if HB is operating in neighborhood c in month t , and β is the coefficient of interest measuring the average effect of HB on shelter entries.

Linear additive-equations like (2) are easier to interpret for policy analysis, but overestimate the effect of HB because many small CDs (small in the sense of shelter entries) started HB late in the period when entries overall were rising. This equation attributes the small absolute size of entry increases in these CDs almost entirely to HB, not to the small size of the entry flow previously.

One way to reduce the bias from estimating (2) is to stratify the sample. The less variation within the dataset in the true counterfactual, the smaller the expected bias from misspecifying the counterfactual. For our data, the geographic variation is much greater than the temporal variation; the Great Recession did not transform the richest neighborhoods of New York City into anything remotely resembling the poorest. Therefore we estimate linear additive models such as (2), but stratify CDs and CTs by geographic linked variation in shelter use.

Mainly, however, we will concentrate on logarithmic equations like (1).

Equations like (1) should be interpreted as reduced form estimates of the average performance of HB during the study period; specifically the geometric mean of the ratio of entries averted to shelter entries.

5.3. Detailed estimating strategy

We will estimate two different sets of equations to find the overall effect of Homebase on shelter entries: the linear and the logarithmic. We estimate both types of equations for both CDs and CTs.

Before we can do that, however, three additional issues must be addressed: families who were served before their CDs were officially operating, foreclosures, and zeroes.

Because of prematurely served families, we treat HB operation as a series of binary variables: H_{ct} , P_{ct} , E_{ct} . H_{ct} is a dummy variable equal to one if and only if HB is officially operating in CD c in month t . To isolate the effect of official from unofficial operations, we

include a second dummy variable P_{ct} , equal to one if and only if some resident of CD c had received HB services during or before month t . We also allow for the possibility that treatment effectiveness may change as time goes on. On the one hand, HB centers might become more proficient as they acquire more experience. In that case, later months of experience would cause greater reductions in shelter entries. On the other hand, HB may delay shelter entries rather than avert them entirely. To explore these possibilities, we included a dummy variable E_{ct} , equal to one if and only if in month t HB has been officially operating in CD c for more than two months. The sum of the coefficients on these three variables estimates the average treatment effect during the period HB was fully operational in a CD. A negative sum indicates that HB reduced shelter entries.

Foreclosures can be seen in two different ways. First, they are an independent variable; we want to know whether foreclosures increase shelter entries. In New York, as in most states, tenants are usually evicted when a multi-family property is foreclosed, and so foreclosures should be seen as shocks that can lead to homelessness. (Landlords facing foreclosure may also unilaterally cease providing services.) Foreclosures are like Homebase in reverse: for any eviction, some of the families would have become homeless anyway, some become homeless because of the eviction, and some would not have become homeless and do not become homeless (or do not become a homeless family, since some foreclosed households are individuals). The proportions should vary by time and neighborhood, and it would not be surprising if the proportion entering homelessness were greater in the times and neighborhoods where more non-foreclosed families were becoming homeless. In this case, foreclosures should be treated just like the binary variables for Homebase, and be added to both the linear and logarithmic equations.

Second, for purposes of estimating HB effects, foreclosures are a useful control variable, and so should improve the estimate of those effects.

Many neighborhood-months have zero shelter entries, and so the logarithmic equations cannot be estimated without making corrections. For CDs, we add one to the number of shelter entries in each CD-month. This will mean that some of our inferences must be slightly reworked when we present results, but the corrections are generally small. For the CTs, we estimate Poisson regressions.

Thus the equations we estimate are the following. For both CDs and CTs, we estimate the same linear equation:

$$s_{ct} = \beta_1 H_{ct} + \beta_2 P_{ct} + \beta_3 E_{ct} + \gamma_c + \delta_t + \sum_s \varphi_s F_{c(t-s)} + \varepsilon_{ct}. \quad (3)$$

For the logarithmic CD equation we estimate:

$$\ln(s_{ct} + 1) = \beta_1 H_{ct} + \beta_2 P_{ct} + \beta_3 E_{ct} + \gamma_c + \delta_t + \sum_s \varphi_s F_{c(t-s)} + \varepsilon_{ct}. \quad (4)$$

For the logarithmic CT equation we estimate:

$$E(\ln s_{ct}) = \beta_1 H_{ct} + \beta_2 P_{ct} + \beta_3 E_{ct} + \gamma_c + \delta_t + \sum_s \varphi_s F_{c(t-s)} + \varepsilon_{ct}. \quad (5)$$

We also experiment with CD-specific time trends.

The model is completed with a series of variables that model the foreclosure process. F_{ct} is a count of the number of units included in initial foreclosure filings in neighborhood c in month t . Because foreclosures take many months to resolve, foreclosure effects are modeled as a set of lag variables extending over 18 months.

All statistical analyses were performed using models in Stata 11.0–13.0.

6. Results

6.1. Logarithmic equations

The logarithmic equations in Table 2 show that Home-base reduced shelter entries by a statistically significant proportion. The CD equation shows that entries would have been 11.2 [95% ci = 3.6, 18.8] log points higher without Homebase when it was fully operational, and the CT equation shows that entries would have been 4.9 [95% ci = -1.0, 11.0] log points higher. The geometric mean CD-month when HB was fully operational had 10.6 augmented entries. Hence average decrease in actual entries in the average CD-month was 11.7%.

We also estimated these equations with CD-specific time trends. Since the HB variables turn on and never turn off, the time trends may over-correct. For CDs, the results implied a slightly smaller HB effect (10.2 log points as opposed to 11.2 log points without CD-specific time trends).

Why the difference between the CT estimate and the CD estimate? A major cause is the way that logarithmic equations aggregate heterogeneous effects. The coefficient of Homebase in a logarithmic equation is the unweighted average percentage decrease that Homebase is associated with across all the units of observation. So the CT coefficient is the unweighted average across all CT-months; that is, it is the unweighted average across CD-months of unweighted averages within CD-months. The CD coefficient is also an unweighted average of percentage changes across within-CD-month averages, but the within-CD-month-averages are not unweighted; instead, the within-CD-month averages are weighted by the share of entries in each CT-month. In other words, the CT coefficients are geometric means both within CD-months and between CD-months, while the CD coefficients are arithmetic means within CD-months and geometric means between CD-months. Everything else being equal, the CD coefficient will be bigger than the CT coefficient if, within CD-months, CT-months with larger percentage reductions in entries because of Homebase also tend to have larger numbers of shelter entries.

6.2. Linear equations

Results of the stratified equations are shown in Table 2. The CD equations imply that Homebase averted 1788 [95% ci = 208, 3367] entries, two-thirds of them from the large stratum. The CT estimates imply 1271 [95% ci = 168, 2, 375] entries averted. In both sets of

results, HB has a bigger effect in the larger strata. For CDs, fully operating HB averts 1.71 entries per CD-month in the large stratum, 0.75 in the small. For CTs, virtually all the entries averted are among families from the moderate-use and high-use CTs. Fully operating HB averted 0.035 and 0.047 entries per CT-month in the moderate and high-use CTs. Since CTs are 32 times more numerous than the number of CDs, the CT-level is comparable to 1.1 and 1.5 entries averted per CD-month. The number of entries per month in the low-use CTs is 0.003 or virtually indistinguishable from zero entries.

6.3. Postponement

Could these results merely reflect shelter entries being postponed, rather than averted? The answer is no. These coefficients reflect average (geometric or arithmetic) impacts of HB over the study period. If all Homebase did was to postpone entries, the estimated impact would be zero, because reduced entries in early months would be offset by increased entries in later months (unless, of course, Homebase managed to postpone entries for several years so that they occurred after the study period). Realize also that Homebase operation in a particular month can have two offsetting effects on future shelter entries. It can increase entries in future months if it postpones entries that would have happened in the current month, or it can reduce entries in future months if it serves today families who would have become homeless in future months.

In all the reported regressions, the coefficient on “experienced” operation is small and insignificant. If either of these two offsetting intertemporal effects were larger than the other on net, this coefficient would have been large in absolute value. Our results suggest therefore that intertemporal effects are small on net (in addition to being accounted for in our results on average).

6.4. Foreclosure effects

Our results on foreclosures generally support statistically significant positive associations between monthly LP filings and the number of shelter entries over the next 18 months in all communities except for high-use CTs. For the CD models, in which we estimated lag coefficients at 3 month intervals, the sum of the foreclosure coefficients is statistically significant but very small in the logarithmic equation: 100 LP listings result in a 0.1% increase in shelter entries over an 18 month period. The results for stratified model imply larger effects and are easier to interpret. For every 100 LP filings in low-use CDs, 3.8 (95% ci = [2.6, 5.0]) additional families enter shelters over an 18 month period; there is no effect in high-use CDs. For the CT models, we estimated all 18 monthly lag coefficients, and the loglinear model indicate that every 100 LP filings is associated with a 1.6% increase in shelter entries over an 18 month period. Turning to the stratified model, 100 LP listings in low-use CTs is associated with 1.3 shelter entries over an 18-month period and 2.3 shelter in moderate-use CTs. Foreclosures in high-use CTs have no effect on shelter entries. The weak effects estimated for the loglinear model may indicate incorrect model specification. The stratified models suggest, contrary to the constant proportional assumption of the loglinear model, a weakening of foreclosure effects in communities with high shelter use.

These results are an economically meaningful effect: the CD coefficient implies that the increase in LP filings in low-use CDs between 2005 and 2008–8304—is associated with an increase of 319 (95% ci = [216, 415]) in shelter entries between those two years; the total increase for these CDs was 844.

Even if they all pointed in the same direction, our data do not allow us to draw a strong causal inference. We cannot state whether this association is a direct consequence of the foreclosure process or whether it is merely an indicator for broader economic factors that are tied to variation in housing stability between neighborhoods and over time. A better understanding of this connection is an important topic for future research.

6.5. Geographic spill overs: “Musical chairs,” contagion, and other effects on non-participants

We do not know whether the shelter entries that HB averted were HB participants or other families; all that we can estimate is the net number of entries averted. But that is the relevant number for evaluating HB, not the number of entries averted among participating families.

HB may either increase or decrease entries among non-participating families. Increases would arise from “musical chairs;” decreases from contagion. Contagion across CD lines could arise, for instance, because many families have friends and relatives who live in other CDs; the example of a family that entered the shelter system might influence other families by giving them information—“how you do it”—or encouragement—“it is n’t that bad”—or reduce stigma—“my friends won’t look down on me because they have been homeless too.” If fewer families are entering shelters, less of this information and encouragement will spread. A priori, there is no way of predicting whether musical chairs or contagion is more powerful.

Do spill overs to non-participants bias our estimate of the net effect of HB? We need to consider each of the several classes of estimate.

For CT estimates, there is no bias. All CTs within a CD are either operating in a given month or not. Spill overs between CTs will cancel each other out in the aggregate. We have estimated the right thing, but we have no information on the auxiliary question of which CT shelter entries were averted.

For CD estimates, a bias could arise if HB activity in an operating CD affected shelter entries in a neighboring CD. If musical chairs were the stronger effect, our estimates of entries averted would be too high; if contagion were the stronger effect, they would be too low.

To check for these geographic spill overs among CDs, we add a new binary variable, “next door” which equals one for a CD-month if and only if any CD adjacent to the CD in question is officially operating. This variable will be positive if operation of Homebase increases entries from neighborhoods “next door,” through “musical chairs,” for instance, or negative if Homebase reduces entries, through contagion.

Table 3 displays the results of equations with the next-door variable, for both the logarithmic equation and the stratified linear equations. The coefficient on next-door is negative and significant in both the logarithmic and stratified linear equations, and it is large enough to be economically meaningful in both equations, too. The simplest story, then, is that spill overs appear to be positive: Homebase operation in a CD-month, far from harming its neighbors, appears to help them.

Further equations not shown here try to determine whether this spillover effect operates mainly on CD-months that are themselves officially operating, or not. Spill over apparently works in both cases, but we are more certain of the effect on CD-months that are not operating. This is because almost all of the operating CD-months next door to operating CDs occur in late 2007 and 2008. For these CD-months we cannot disentangle the effect of being next-door to an operating CD from several other reasons that might explain greater efficacy in those years, such as learning or a greater concentration of Homebase in Manhattan and Queens.

Thus we are pretty sure that Homebase did not hurt non-participants on net, and there is good reason to think that it helped non-participants across CD lines, especially in small-stratum CDs. So our main estimates may be biased downward on this ground.

Since we cannot observe whether Homebase helped or hurt non-participants within CD borders, we should expect that it helped. Of course, whatever these effects are within CT or CD boundaries, they are already accounted for in our estimates.

7. Policy refinements

Homebase operated in different ways in different times and places. The policy refinement equations are one way to try to assess the impact of different ways of operating Homebase, and thereby make the process more transparent. The idea is to divide the operating neighborhood-months into several different classes, and estimate for each class a function of entries averted per neighborhood-month. For ease of exposition, assume the classes form a partition of the operating neighborhood-months (this assumption is inessential). Let $j = 1, 2, \dots, J$ denote the different classes, and take class J as the reference class. Let K_{ctj} be a dummy variable equal to one if and only if neighborhood-month (c, t) is in class j .

Then the policy refinement variant of Eq. (4) is

$$\ln(s_{ct}+1) = \beta_1 H_{ct} + \sum_{j=1}^{j-1} \lambda_j K_{ctj} + \gamma_c + \delta_t + \sum_s \varphi_s F_{c(t-s)} + \varepsilon_{ct}. \quad (5')$$

Then λ_j will be an unbiased estimate of the average log decrease in entries experienced in class j relative to class J . We will learn which classes of operating months produce larger or smaller relative decreases in entries.

Of course, these results are only suggestive. Policies were not chosen randomly, and their implementation may coincide with other events happening over time or space. We do not have a great deal of independent variation in policies.

We work with several different partitions of the CD-months: by shelter-use stratum, by season (summer vs non-summer), by location of HB centers, and by cohort. The results are presented in Tables 4a and 4b. For each partition, we test to see whether we can reject the hypothesis that all coefficients associated with official operations are the same. (In terms of Eq. (5), Tables 4a and 4b display $(\beta_1 + \lambda_j)$ when j is not the excluded condition, and β_1 when it is. Thus these tables show the rate at which official operation of Homebase averts entries when the condition applies. We include the “unofficial operation” variable, but omit the “experienced” variable for simplicity.)

7.1. By stratum

For CDs, Homebase averted a statistically significantly higher proportion of entries in the small use than it did in the large stratum (Table 4a). Hence we can reject the null hypothesis of homogeneous effects. Official Homebase operations reduced entries by 16.6 log points in the small-use CD stratum, 5.9 log points in the large-use stratum. But the large stratum had more entries, of course. The CT-level results were the reverse. Official Homebase operations were associated with an increase in shelter entries in the low-use stratum (where entries are rare), whereas it was associated with 8.2 and 11.7 log point reductions in entries in moderate-use and high-use CTs respectively.

7.2. By season

We used two seasons—summer (July and August) and everything else—because DHS believed that shelter entries were higher in the summer, and because our preliminary look at the data confirmed this. For both CDs and CTs, we could not reject the null hypothesis that Homebase averted the same proportion of entries in the summer as in the rest of the year (Table 4a).

7.3. By location of Homebase centers

Not all operating CDs have Homebase centers located in them; in many cases, eligible families must travel to a different neighborhood in order to find a center that will serve them. Families in CDs without their own Homebase centers may have less information about the program, and traveling to a center outside the neighborhood is expensive and time-consuming. So the number and mix of families that go to Homebase centers from close CDs should be different from the number and mix from far CDs.

For entry reductions, several offsetting effects could be operating. If families who overcome more informational and transportation hurdles to arrive at Homebase are more likely to have problems that would soon place them in shelters and are more highly motivated to solve those problems, then the proportion of Homebase participants whose shelter entry is averted should rise. We call this the selection effect. On the other hand, if fewer families arrive at Homebase centers because of information or transportation problems, Homebase will have fewer opportunities to avert entries. We call this the volume effect. Finally, Homebase

centers were not randomly placed; they may have been located in neighborhoods where the most entries could be averted, and so in general we should expect more entries to be averted from these neighborhoods. We call this the design effect.

To see if what the net effect was, we partitioned the officially operating CD-months into three sets: those in which a Homebase center was present; those adjacent to a CD with a Homebase center; and all other officially operating CD-months (“isolated”). If the selection effect is dominant, more entries would be averted from adjacent and isolated CD-months than from the CD-months in which a Homebase center was operating; if the volume and design effects are dominant, the opposite will occur.

Table 4b shows the estimated coefficients in the policy refinement equation for center location. Table 5 shows basic information about the operation of Homebase during the different kinds of CD-months.

Note first from Table 5, that the necessary conditions for the volume and selection effects seem to be met. HB centers appear to have been placed in CDs with a large number of shelter entries (although timing also plays some role, since in the later years when more families were entering shelters, more Homebase centers were open). HB centers also appear to draw more families from the immediate neighborhood than from more distant neighborhoods (although timing plays a role here, too, since the pre-2007 Big Six CDs had individual centers and unusually high volumes).

For CDs, the coefficients in Table 4b indicate that HB averted a larger proportion of entries from adjacent and especially isolated CDs than from nearby CDs. Having an operating center reduced the proportion of entries averted, rather than increasing it. This is strong support for the selection effect. For CTs, the coefficients are effectively identical: having an operating center does not increase the proportion of entries averted. The same general pattern appears to hold with CTs, but not so strongly. This does not indicate a rejection of the selection effect, since the weakest statement of that effect is that entries averted per Homebase family should be lower for CD-months in which a center was operating.

7.4. By cohort

The Big Six cohort operated for several years with very large caseloads, and before a great deal of knowledge could be accumulated about how to operate Homebase. The large caseloads could have two opposing effects, as we outlined in the discussion of center locations. The volume effect—more families whose possible homelessness could be averted—would increase the number of entries averted in these CD-months, and selection effect—a higher proportion of families who would not otherwise be homeless or whose problems Homebase would be unable to alleviate—would decrease entries averted. So we can learn something by comparing the early years of the Big Six with later years of the Big Six and with the two other cohorts.

The patterns for the CDs and the CTs are about the same: a high proportion of entries were averted in the 2007 cohort, and a small proportion were averted in the Big Six cohort; in fact, after 2007, the Big Six may have increased shelter entries rather than decreased them.

The CD and CT models send conflicting messages about the 2008 cohort, but both the CD and CT models indicate that the rate of averting shelter entries was not significantly different from zero for this cohort. The selection effect seems to be important: the large volume of service during the pre-2007 days of the Big Six did not result in a large number of entries averted, although the pre-2007 Big Six appear to be more successful than the post-2007 Big Six. (Learning would also imply that the post-2007 Big Six would be more successful than the pre-2007 Big Six.) We do not know why the 2007 cohort appears to be so successful; if learning and the passage of time were the whole story, we would expect the 2008 cohort to be even more successful, but it is not. The different cohorts were concentrated in different boroughs, and so we might be picking up a borough effect rather than a cohort effect.

We also conducted an F -test of the null hypothesis that the effect of official operation of Homebase was the same across all cohorts. The test rejected this hypothesis at the 0.0001 level.

8. How many entries were averted?

How many entries Homebase averted is a natural question to ask. As we have mentioned, the linear CD stratified equations imply that Homebase averted 1788 entries (95% ci = [208, 3367]) and the linear CT stratified equations imply somewhat smaller number of entries averted, 1271 (95% ci = [168, 2,375]), and there is a substantial overlap in the 95% C.I. intervals. However, our preferred specification, the logarithmic, does not answer this question. The logarithmic results tell us what the average percentage reduction in entries was in a typical CD- or CT-month, but that does not imply any particular arithmetic reduction in entries.

To see why, consider two scenarios with the same average percentage reduction in entries. In the “positive correlation scenario,” neighborhood-months with large counterfactual shelter entries have large percentage reductions and neighborhood-months with small counterfactual shelter entries have small percentage reductions. The “negative correlation scenario” is the opposite. Obviously the overall absolute reduction in entries is greater in the positive correlation scenario than in the negative, even though the average percentage reduction (what the logarithmic equations estimate) is the same.

What is the range of entries averted that could be implied by any particular percentage reduction? In the Appendix, we show that this range has a lower bound, but no upper bound.

The lower bound is achieved when we assume that the counterfactual number of entries in the absence of Home-base would have been the same in every neighborhood-month. That constant counterfactual number of entries is the geometric mean of the actual number of entries, times $\exp(-\beta)$, where β is the coefficient of Homebase in the logarithmic equation. In the lower bound case, counterfactual entries from every neighborhood-month are the same as the counterfactual entries from the geometric mean neighborhood-month, if Homebase had its average effect in that month. We can think of the lower bound estimate as arising from the case when counterfactual entries are constant.

In our data, the lower-bound calculation always yields a negative number of entries averted. This is because the geometric mean of actual entries is considerably below the arithmetic mean (for instance, for fully operating CD-months, the arithmetic mean is approximately 16.6 and geometric mean is approximately 9.6; for fully operating CT-months the arithmetic mean is 0.50 and the geometric mean is 0.083). If the counterfactual entries are 10–20% greater than the geometric mean, then they are still smaller on average than the arithmetic mean of actual entries. Since with actual entries there are many CD-months with few entries and a few CD-months with many entries, the counterfactual of every month with, say, 10% more entries than the actual geometric, implies that Homebase caused many small decreases and a few large increases, and the large increases outweighed the small decreases. That made the geometric mean go down and the arithmetic mean go up.

9. Comparison with the randomized control trial

As we have mentioned, Rolston et al. (2013) reported on a small controlled experiment. The control group received advice, and was eligible for other non-Homebase services. They found that among families offered HB services, applications for shelter were lower by 8.9% (95% ci estimated as [17.2, 0.7]) and shelter entries by 6.5% (95% ci estimated as [14.0, -1.2]). Thus they also find that Homebase probably reduced shelter entries. The size of the effect they find, while not strictly comparable to our work because they estimate entries averted per family served, seems to fall between the CD and CT results, and could easily be the effect for one of the conditions or time periods we study in Section 7. How do our results differ from theirs, then?

The most unimportant difference is that Rolston et al. excluded adult families from their sample.

The main differences are generalizability and heterogeneity. Rolston et al. could look only at one version of Homebase operating only in one particular environment—families who applied between June and September 2010. Their “effect” of Homebase is by necessity a ratio of entries averted to families served. Both theory and empirics tell us that this ratio is not a constant, and so the response that Rolston et al. found in June–September 2010 tells us little about what the program effect was or will be in other times and circumstances. As Remler and Van Ryzin (2014, chapter 14) point out, “Although randomized experiments provide convincing evidence of cause and effect, this often comes at a price: limited generalizability.” Generalizability is a particularly serious problem when the effects being measured are heterogeneous. Rolston et al. studied a particular moment in the history of Homebase—summer, right after the Great Recession, Homebase operations constrained by federal operations because of a short period of federal funding, and the shelter system itself in transition to much tighter rules on post-shelter subsidies. They cannot say whether that moment was representative.

But we can. Our contribution is to affirm that the moment that Rolston et al. studied was reasonably representative of the early years of Homebase—but they could easily have picked a moment that was unrepresentative. Heterogeneity is a hurdle for a quasi-experiment, but a

road-block for an RCT. Heterogeneity has been the greatest difficulty we faced in writing this paper, but ultimately heterogeneity is what makes it worth reading.

We also differ from Rolston et al. in looking at how Homebase affected non-participants as well as participants. (Remler and Van Ryzin label these issues “general equilibrium effects.”) The major recent controversy about New York City family homelessness—whether and to what extent the offer of post-shelter rental subsidies induces families to enter shelters (Cragg and O’Flaherty, 1999; O’Flaherty and Wu, 2006; NYC Independent Budget Office, 2012)—hinges on how a program affects non-participants. Evaluations of New York City homelessness programs should not ignore non-participants.

The controlled experiment also provides no way of looking at the process by which families arrive at Home-base centers—possibly the most important part of the Homebase process. They study only the families who actually arrived at centers in the summer of 2010, not the families who did not arrive or might arrive under other circumstances. In their study, the ratio of entries averted to actual entries is about one, far greater than anything we saw, and far greater than any plausible number that could be attached to actual data (if the ratio were about one in the period we study, Homebase would have averted at least 16,000 entries). While we cannot tell much about the crucial recruitment process, on which any expansion of Homebase depends, we can at least observe aspects of it.²

In summary, we think that this paper complements Rolston et al. (2013).

10. Discussion

Our results indicate that in neighborhoods in which Homebase was operating, 5–11% of the families who would have otherwise entered shelters did not do so. The most intriguing question is why the other 89–95% of families of counterfactual shelter entries were not also averted.

Some families whom Homebase helped entered shelters nevertheless, but Rolston et al. (2013) and Shinn et al. (2013) both find that the number of treatment failures is small, about a tenth of the families that Homebase serves, and the number of families whom Homebase serves is usually less than the number of shelter entries. Thus even when Homebase is fully operating, the overwhelming majority of families who enter shelters have never visited Homebase. A major question for future research is why.

Another way of phrasing this question to ask what makes the families who decide on their own to visit Homebase different. Is it information about Homebase? Information about the housing problems they are experiencing? Are the problems they are experiencing different from the problems of shelter entrants who avoid Homebase, or are they just more aware of

²Another difference is that the RCT can shed some light on the question of what happens to families who might be turned away from Home-base; we can say almost nothing about this. Aside from being less likely to enter shelters and less likely to stay in shelters for a long time, there were no significant differences between the control and experimental groups in any of the measures that the investigators looked at: child protective system involvement, receipt of income support payments, and employment. Since child protective services in New York City would be highly likely to remove the children from any family that was sleeping on the streets, these results suggest strongly that control families did not end up sleeping on the streets.

them? Are they motivated differently? Are they more desperate? Or are they just lucky in some way to stumble across Homebase?

The families who visit Homebase centers are unusual not only among families who would enter shelters, they are also unusual among all families in New York City. In the average year we studied, about one New York City family (household with more than one person) out of 200 entered the shelter system, but in the control group of Rolston et al. (2013) about one out of seven entered the shelter system. How did the families who visited Homebase centers decide that they were different from the rest of the community?

The families who walk through its doors are the input that it would be hardest for Homebase to find a substitute for, and we have no idea how they get there. The other inputs—the services that Homebase provides—also matter, of course, and they should be studied too, but even the efficacy of services cannot be considered separate from their effects on who walks through the door.

Understanding what brings families through Homebase's doors is also important for assessing the prospects for expanding or replicating Homebase. Our results may not apply beyond New York, where shelter costs are unusually high because of a court-mandated right-to-shelter and a large family shelter system, in which families stay for long periods.³ On the other hand, New York is unusual also in its safety-net programs outside the homeless system. Homebase operated in an environment where the public assistance agency had a well-developed program for giving emergency assistance—“one-shot” aid—and also had legal responsibilities for emergency housing assistance (“Jiggets aid” and later the Family Eviction Prevention Supplement). The budget for these programs in 2008 was an order of magnitude greater than the Homebase budget (NYC Independent Budget Office, 2008). The Administration for Children's Services, the child protection agency, also spends several million dollars a year to help prevent foster care placements that would occur because of inadequate housing and homelessness; in 2008 almost a thousand families were assisted (NYC Independent Budget Office, 2008). To our knowledge, there is no evidence that the intensity or scale of these programs varied by geography in ways that would bias our estimates of the effectiveness of Homebase. Instead, we might view the existence of such programs as complementary, if part of Homebase's services is to refer eligible families to other programs through which they could gain temporary financial and legal support. While there is anecdotal evidence from employees at DHS that such referrals were in fact part of the multifaceted services that were offered to clients, Rolston et al. (2013) do not find evidence of increased payments made through one-shot assistance programs.⁴ Nonetheless, since we are agnostic about how exactly Homebase is reducing shelter entry, differential take-up of additional assistance programs among Homebase families would not alter the size of our coefficients on families helped. In fact, what we are measuring is the added effect of

³Nationally, the median shelter stay for a family in 2008 lasted 30 days (U.S. Department of Housing and Urban Development, 2009, p. 37), while the average stay in New York was about a year. In the U.S. as a whole, only New York City, Massachusetts, and Washington DC have had legal mandates that might be interpreted as “right-to-shelter.” See Main (2016).

⁴Rolston et al. (2013) compare only the number of payments made, not the level of spending, through one-shot assistance between their treatment and control groups.

Homebase above one-shot aid and the Family Eviction Prevention Supplement. In locations without such programs, Home-base might have even larger effects.

Local housing and homeless policy also matter in determining prevention program effectiveness. Homelessness can be prevented only among people who would otherwise be homeless, and local housing policies and the structure of homeless programs together exert considerable influence on who homeless people are. In the limit, homelessness prevention would not “work” at all in a city with no shelters (and Draconian street policies). More expensive and attractive shelters may make homelessness prevention a more attractive strategy from the point of view of a municipal budget, but a somewhat less attractive strategy from the point of view of family well-being.

We conclude with a simple cost-benefit analysis of Homebase, but note a couple of caveats regarding its interpretation. First, we include as benefits only the reduction in shelter operating costs associated with our Homebase aversion estimate, which likely understate the true benefits of the program, both to individuals and society. We do not have a way of valuing the primary goals of Homebase: the benefits to the participants themselves of both the direct assistance that Homebase provided (even if it did not avert homelessness) and the benefits both to participants and the rest of society from averting homelessness. Home-base may have promoted residential stability, for instance, quite apart from any effect on homelessness, and a large body of evidence (Aaronson, 2000; Astone and McLanahan, 1994; Haveman et al., 1991; Mohanty and Raut, 2009) suggests that residential instability hurts children's cognitive development. Because shelter stays in New York are long and expensive, the benefits to city and state taxpayers alone may have been as large as the costs. Separately, our cost estimate too, which includes only the Homebase operating budget, is likely understated, especially if, as noted above, a critical feature of Homebase is the referral of eligible families to the other New York public assistance programs. The cost of averting shelter entries is better approximated by the total spending of Homebase together with the spending increase, if any, driven by Homebase referrals in these programs.

In city fiscal year 2009, the last year for which we have any data and the first year during which Homebase operated citywide, expenditure for family homeless shelters was \$400 million (NYC Independent Budget Office, 2015, p. 9) and the budget for Homebase was \$14.2 million (NYC Independent Budget Office, 2008, p. 3). A reduction of 5–11% in shelter entries should reduce shelter population in the long run by 5–11%, since Homebase does not noticeably alter shelter exit hazards (Goodman et al., 2014). If the shelter budget is roughly proportional to shelter population, the \$14.2 million Homebase budget would have reduced shelter expenditures by \$20–44 million.⁵

Community-based prevention programs may prove to be an important policy tool in reducing family homelessness. This study is a step in a larger program of research on how they work in New York City and how they might be adopted elsewhere.

⁵To be sure, the family shelter budget includes expenditures for families without children, but Homebase also serves both families without children and single adults from its budget.

Acknowledgments

We are grateful to the New York City Department of Homeless Services for data, financial assistance, and answers to our many questions; especially to Jay Bainbridge, Joanna Weissman, Eileen Lynch Johns, Sara Zuiderveen, Jonathan Kwon, Veronica Neville, Maryanne Schretzman, and Ellen Howard-Cooper. Scott Auwater and Bronxworks provided us with a detailed look at an operating program. We are grateful for support from our institutional partners at CUNY, John Mollenkopf and Ellen Munley; and from the Columbia Center for Homelessness Prevention Studies, especially Carol Caton, Bill McAllister, Sue Marcus, Sue Barrow, Mireille Valbrun and Shoshana Vasheetz. We have received helpful advice from Serena Ng, Kathy O'Regan, Christoph Rothe, Bernard Salanie, Beth Shinn and Till von Wachter; and participants at conferences sponsored by the National Alliance to End Homelessness, NYC Real Estate Economics Group, NYU Furman Center and the Columbia Population Research Center. Maiko Yomogida, Abhishek Joshi provided excellent research assistance. Financial assistance from the New York City Department of Homeless Services, the National Institute of Mental Health (5 P30MH071430-03), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (1R24D058486) is gratefully acknowledged. The errors are our own. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health, the Federal Reserve Board of Governors, or the Department of Homeless Services.

Appendix

A.1. Part one—theory of homelessness prevention

Suppose that families can be arrayed along a single dimension of housing distress z , with larger values of z corresponding to greater distress. Without Homebase assistance, families with $z > Z_1$ will become homeless very soon, but with Homebase assistance, only families with $z > Z_2$ will become homeless. Homebase draws its clientele from families with $z > Z_0$ (where $Z_0 < Z_1 < Z_2$), but otherwise it cannot observe z . (Both Shinn et al. (2013) and Rolston et al. (2013) suggest that over half of the families that Homebase serves would not have become homeless in the absence of Homebase. Thus we seem safe in maintaining that $Z_0 < Z_1$.) Homebase centers serve every family with $z > Z_0$.⁶

Let $F^*(\cdot|\mu)$ denote the cdf of housing distress in the general population of this neighborhood in this month. We assume that this distribution is indexed by a single parameter μ that moves the distribution but does not deform it:

⁶Shinn et al. (2013) in their study of Homebase raise questions about whether this “Goldilocks” view of homelessness prevention is accurate. Using characteristics recorded on intake forms, they calculate an index which is monotonically related to the probability of shelter entry (that is, treatment failure) both for Homebase participants and for families that applied for Homebase and were found ineligible. The difference between the shelter entry rate for ineligibles and the shelter entry rate for eligibles is a monotonically increasing function of this index. Let w denote the Shinn index. Clearly, if two families have the same Shinn index but one is ineligible and the other is not, there is some other difference between the families. Let $G_I(z|w)$ denote the cdf of housing distress for an ineligible family with Shinn index w , and $G_E(z|w)$ the cdf for an eligible family. Then the result is that

$$(1 - G_I(Z_1|w)) - (1 - G_E(Z_2|w)) = G_E(Z_2|w) - G_I(Z_1|w)$$

is increasing in w , while both $G_E(Z_2|w)$ and $G_I(Z_1|w)$ are decreasing in w . Many pairs of cdf's satisfy this restriction. For instance, suppose that for eligible families, z is distributed uniformly on the range $[w - 1, w + 1]$ given w , and for ineligible families, is distributed uniformly on the range $[kw - 1, kw + 1]$ with $k > 1$. Ineligibles have greater housing distress (worse attitude, for instance) than eligibles, as indicated by information that Homebase sees but does not record for use of econometricians. Then the

probability of shelter entry for eligibles is $\frac{1}{2}(w+1 - Z_2)$ and the probability of shelter entry for ineligibles is $\frac{1}{2}(kw+1 - Z_1)$ for $w \in [\min(Z_2 - 1, \frac{Z_1 - 1}{k}), \frac{Z_1 + 1}{k}]$. (The restriction on w assures that both probabilities are strictly between 0 and 1, as are all the probabilities that Shinn et al (2013) report.) Then the probability of shelter entry rises faster as w increases for all observed w , even though in fact the Goldilocks condition holds.

$$F^*(z|\mu) = F^*(z+x|\mu+x) \text{ for all } z, \mu, x.$$

This variable μ includes such influences as general macroeconomic conditions, housing market conditions, and DHS policies.

For each z let $p(z)$ denote the proportion of the population with housing distress z who arrive at Homepage. Let

$$f(z|\mu) = p(z) dF^*(z|\mu)$$

and

$$F(z|\mu) = \int_{-\infty}^z f(z'|\mu) dz'.$$

Then $F(\cdot|\mu)$ is cumulative distribution of the population who seek out Homepage. It is easy to show that this distribution, like its parent distribution, is indexed by the single location parameter μ that moves the entire distribution but does not deform it:

$$F(z|\mu) = F(z+x|\mu+x) \text{ for all } z, \mu, x.$$

In particular then,

$$\frac{\partial F(z|\mu)}{\partial \mu} = -f(z|\mu).$$

Let $P(\mu)$ denote the population who seek out Homepage

$$P(\mu) = \int_{-\infty}^{\infty} f(z'|\mu) dz'.$$

Then the number of families that Homepage serves is $P(\mu) - F(Z_0|\mu)$, the number of Homepage families who would become homeless without Homepage is $P(\mu) - F(Z_1|\mu)$, the number of families who become homeless when Homepage is operating (the treatment failures) is $P(\mu) - F(Z_2|\mu)$, and the number of entries that Homepage averts is $(F(Z_2|\mu) - F(Z_1|\mu))$. The number of families who enter the shelter system is the sum of the Homepage treatment failures and the families above the untreated homelessness threshold who did not visit Homepage:

$$\int_{Z_1}^{\infty} dF^*(z') - (F(Z_2|\mu) - F(Z_1|\mu)) = \int_{Z_2}^{\infty} dF^*(z'|\mu) + \int_{Z_1}^{Z_2} (1 - p(z')) dF^*(z'|\mu).$$

Counterfactual entries—the number of families who would enter the shelter system in the absence of Home-base is simply $\int_{Z_1}^{\infty} dF^*(z'|\mu)$.

A.1.1. Non-administrative variation

Variation among months and neighborhoods not caused by DHS or Homebase operators is caused by variation in μ . Let μ increase. Housing market distress becomes greater (and assume no administrative response for now). The change in counterfactual homelessness is $dF_*(Z_1|\mu) > 0$. The first question is whether greater counterfactual homelessness is associated with more averted homelessness or less. The change in averted homelessness is

$$\frac{\partial}{\partial \mu} (F(Z_2|\mu) - F(Z_1|\mu)) = f(Z_1|\mu) - f(Z_2|\mu)$$

Thus if $f(\cdot)$ is falling in the range where Homebase averts entries, more counterfactual homelessness coincides with more entries averted; if $f(\cdot)$ is flat, entries averted are unchanged; and if $f(\cdot)$ is rising, more counterfactual homelessness coincides with fewer entries averted.

A priori we cannot say whether the pdf $f(\cdot)$ is rising, flat, or falling in the relevant range. If the population distribution $F_*(\cdot)$ is unimodal (normal, for instance), and the relevant population for Homebase has greater than modal distress, then the pdf of $F_*(\cdot)$ will be decreasing in the relevant range. But $p(\cdot)$ might be rising (for instance, if families who think they are more likely to derive benefit are more likely to visit Homebase, and if they are reasonably well informed). So no definitive statement about $f(\cdot)$ can be made.

We begin by examining two special cases, and then look at the general case. For the most part, we will assume that DHS does not actively respond to changes in housing distress, but instead accommodates them. That is an expositional tool. In the next section, we will investigate what happens when DHS takes active steps, either because of changes in housing distress or for some other reason. If DHS always responds to changes in housing distress in a particular way, then the data will show just the compound effect of the housing distress change and the administrative change.

In the first special case, the distribution of housing distress is a displaced exponential, and changes in the environment move the location of this distribution. For this special case, everything is proportional: the ratio of entries averted to actual entries is a constant; the ratio of entries averted to families served is a constant (as is often assumed in practical discussions); and so the ratio of families served to actual entries is a constant, too. The last proposition is empirically testable, and our data reject this proposition overwhelmingly.

In the second special case, the distribution of housing distress is uniform over the relevant range. In this case, the number of entries averted per neighborhood month is a constant, but the number of families served is not. So the ratio between these two numbers is not a constant.

A.1.2. Special case: displaced exponential distribution

Suppose that $p = p(z)$ is a constant and that $F_*(\cdot)$ in any neighborhood-month is a displaced exponential distribution whose location varies by neighborhood-month:

$$F^*(z|\mu_{ct}) = 1 - \exp[-\lambda(z - \mu_{ct})], z \geq \mu_{ct}.$$

Then the distribution of families whom Homebase sees is:

$$F(z|\mu_{ct}) = p - p \exp[-\lambda(z - \mu_{ct})], z \geq \mu_{ct}.$$

Let s_{ct} denote the actual number of shelter entries from neighborhood c in month t , and let S_{ct} denote the counterfactual number—the number that would have entered if Homebase had not been operating. Then

$$s_{ct} = \exp[-\lambda(Z_1 - \mu_{ct})] - p \{ \exp[-\lambda(Z_1 - \mu_{ct})] - \exp[-\lambda(Z_2 - \mu_{ct})] \} = \exp[\lambda\mu_{ct}] \{ (1-p) \exp[-\lambda Z_1] + p \exp[-\lambda Z_2] \}$$

$$S_{ct} = \exp[-\lambda(Z_1 - \mu_{ct})] = \exp[\lambda\mu_{ct}] \exp(-\lambda Z_1)$$

and so the number of entries averted is

$$S_{ct} - s_{ct} = p \exp[\lambda\mu_{ct}] \{ \exp[-\lambda Z_1] - \exp[-\lambda Z_2] \}.$$

Thus the ratio between number of entries averted and actual entries is a constant:

$$\frac{S_{ct} - s_{ct}}{s_{ct}} = p \frac{\exp[-\lambda Z_1] - \exp[-\lambda Z_2]}{(1-p) \exp[-\lambda Z_1] + p \exp[-\lambda Z_2]}.$$

Moreover, the number of families served in a neighborhood month is also in a constant ratio with the number of entries averted (and so also in a constant ratio with the number of actual entries):

$$H_{ct} = p \exp[-\lambda(Z_0 - \mu_{ct})] = \exp[\lambda\mu_{ct}] \exp[-\lambda Z_0]$$

$$\frac{H_{ct}}{S_{ct} - s_{ct}} = \frac{\exp[-\lambda Z_0]}{\exp[-\lambda Z_1] - \exp[-\lambda Z_2]}$$

$$\frac{H_{ct}}{s_{ct}} = \frac{p \exp[-\lambda Z_0]}{(1-p) \exp[-\lambda Z_1] + p \exp[-\lambda Z_2]}$$

The latter proposition, the constant ratio between entries and families served, is testable; in our data it fails spectacularly. Thus we are pretty sure that this case, when entries averted are a constant fraction of families served, does not obtain in our data. Entries averted per family served are not a constant.

A.1.3. Special case: uniform distribution

Suppose $F_*(\cdot)$ is uniform in the relevant range; that $p(\cdot)$ is a constant; and so $F(\cdot)$ also uniform in the relevant range.

Let ψ denote the constant density of the population distribution and $Z^* + \mu_{ct}$ the greatest value with positive probability. Then

$$\begin{aligned} S_{ct} &= \psi[Z^* + \mu_{ct} - Z_1] = \psi[Z^* - Z_1] + \psi\mu_{ct} \\ S_{ct} - s_{ct} &= p\psi[Z_2 - Z_1] \\ s_{ct} &= \psi[Z^* - (1 - p)Z_1 - Z_2] + \psi\mu_{ct}. \end{aligned}$$

Thus the number of entries averted is a constant. The number of families served is

$$H_{ct} = p\psi[Z^* - Z_0] + p\psi\mu_{ct}.$$

Since the number of families served varies and the number of entries averted does not, the ratio between these two numbers is not a constant.

A.1.4. General case

In general, let s_{ct} denote the actual number of shelter entries from neighborhood c in month t , and let S_{ct} denote the counterfactual number—the number that would have entered if Homebase had not been operating.

When μ increases, the change in the number of entries averted, as we have seen is

$$\frac{\partial}{\partial \mu}(F(Z_2|\mu) - F(Z_1|\mu)) = f(Z_1|\mu) - f(Z_2|\mu)$$

This could be either positive, negative or zero. For the observable quantities, if DHS accommodates the change:

$$\begin{aligned} \frac{\partial}{\partial \mu} H_{ct} &= f(Z_0|\mu_{ct}) \\ \frac{\partial}{\partial \mu} s_{ct} &= f(Z_1|\mu_{ct}) \left[\frac{1}{p(Z_1)} - 1 \right] + f(Z_2|\mu_{ct}). \end{aligned}$$

Thus in general, none of these quantities are proportional to each other, when housing distress varies. There is no stable ratio between entries averted and either families served or actual entries.

What happens if Homebase cannot observe whether $z \geq Z_0$? Suppose instead that it observes some variable w that is correlated with z , and serves all households with $w \geq W_0$. To be specific, let z and w be distributed bivariate normal with the positive correlation ρ : $(z, w) \sim N(\mu, a_w + bw\mu; 1, 1, \rho)$ where $b_w \geq 0$. Then the number of entries averted is

$$A = \int_{W_0}^{\infty} \int_{Z_1}^{Z_2} \varphi(z - \mu, w - a_w - b_w\mu | 1, 1; 1, 1; \rho) dz dw.$$

Differentiate with respect to μ to obtain:

$$\frac{\partial A}{\partial \mu} = \int_{W_0}^{\infty} \int_{Z_1}^{Z_2} \left[-\frac{\partial \varphi}{\partial z} - b_w \frac{\partial \varphi}{\partial w} \right] dz dw.$$

If, as is likely, both z and w are above their respective modes in the relevant range, then both

$\frac{\partial \varphi}{\partial z}$ and $\frac{\partial \varphi}{\partial w}$ are negative and $\frac{\partial A}{\partial \mu}$ is positive. By the same reasoning, greater housing market distress will cause more averted entries even if the pdf of z is flat as long as the pdf of w is downward sloping. But there is no stable ratio between entries averted and either families served or actual entries.

A.1.5. Administrative variation

Homebase activity can also change because DHS or Homebase operators change the program—either in the absence of changes in housing distress or as a response to them.

Administrative decisions are reflected in this formulation in several ways. First, administrators can vary Z_0 , the eligibility threshold. In the limit, if Homebase is not operating, $Z_0 = \infty$. Second, administrators can increase the resources available for each case, and so raise or lower the threshold for treatment failure Z_2 . Third, administrators can vary outreach and publicity efforts, and change centers' locations and operating hours to make them more or less accessible. This will change the function $p(\cdot)$ and hence the function $F(\cdot)$. We do not know how often, where, or when DHS and Homebase operators used each mode of adjustment.

How the outcomes change depends on what aspect of the program varies.

First, consider changes in Z_0 : DHS varies the desired number of families served by varying the eligibility standard. As long as $Z_0 < Z_1$, the number of entries averted remains unchanged. Only the number of families served changes. The ratio between families served and entries averted changes.

Next, consider changes in Z_2 : DHS varies the resources available per case and so the threshold for treatment failure changes. Then the number of entries averted changes, but the number of Homebase families served stays the same. Once again, the ratio changes.

Finally, consider changes in outreach and recruitment. These change the distribution of housing distress among the families appearing at Homebase centers. Many different stories can be told. If $f(\cdot)$ shifts uniformly, then entries averted and families served move in the same direction. Only in this case will the two numbers move in the same proportion. On the other hand, if before the change DHS was operating an optimal outreach and recruiting plan, then any expansion should be subject to (weakly) diminishing marginal returns. Entries averted will rise, but the ratio to Homebase families served will fall.

What about the relationship between counterfactual entries and actual entries when Homebase policies change? Since Z_0 affects neither value, resource changes that operate through eligibility changes do not affect this relationship. If policies change resources per case and so cause Z_2 to vary, actual entries change but counterfactual entries do not. Only when the policy changes operate through outreach and recruitment might counterfactual entries and actual entries move in the same direction.

A.2. Part two—functional form issues

The difference-in-differences methodology assumes that the unobserved S_{it} depends on a neighborhood-specific fixed effect α_i and a month-specific effect β_t . However, the functional form of this relationship is not specified. Two obvious candidates for functional form are the additive

$$S_{it} = \alpha_i + \beta_t$$

and the multiplicative

$$S_{it} = \alpha_i \beta_t$$

The question for our data set is whether it is more realistic to think of the Great Recession (and changes in system-wide DHS policies like the Advantage program) as causing equal absolute increases across neighborhoods in entries (the additive case) or equal relative increases (the multiplicative case).

Both intuition and the data suggest the multiplicative case. In this section, assume that the counterfactual number of entries in each neighborhood-month is observable.

We might estimate the average effect of Homebase on shelter entries in the neighborhood-months in which it was operating with the equation

$$s_{ct} = S_{ct} - H_{ct}\beta \quad (A1)$$

where H_{ct} is a dummy variable equal to one if and only if Homebase is operating in neighborhood-month (c,t) . If (A1) is estimated by OLS, then β is an unbiased estimate of the average decrease in entries associated with Home-base operation, no matter what the relationship is between entries averted, counterfactual entries, and number of families served. Note that (A1) is not a structural equation; it is a way of finding the arithmetic difference between counterfactual and average entries.

We might also estimate

$$\ln s_{ct} = \ln S_{ct} - H_{ct} \tilde{\beta} \quad (\text{A2})$$

by OLS. Then $\tilde{\beta}$ is an unbiased estimate of the average decrease in the natural log of entries associated with Homebase operations, and from $\tilde{\beta}$ we can construct an unbiased estimate of the geometric mean of counterfactual entries. Like (A1), (A2) is not a structural equation, and this geometric mean property holds whether or not the number of averted entries depends on S_{ct} .

The only difference between (A1) and (A2) is in the type of mean that they produce, and so they are equally valid ways of expressing the effect of Homebase on shelter entries.

Of course, we cannot observe counterfactual entries; we must estimate them. Since we have demonstrated above that the multiplicative form is probably the best way to approximate counterfactual entries, the difference-indifferences (DD) equation we would like to estimate is

$$s_{it} = AH_{it} + \sum_i B_i 1_i \sum_t C_t 1_t \quad (\text{A3})$$

The difficulty is that standard methods do not estimate (A3). We do not know how to estimate it. Moreover, if we tried to estimate it, our results would not be comparable with any existing studies.

There are two alternatives. One possibility is to estimate the logarithmic equation

$$\ln s_{it} = AH_{it} + \sum_i B_i 1_i + \sum_t C_t 1_t. \quad (\text{A4})$$

This is essentially (A2), and so will give us interesting information about geometric means, but if we want to estimate total entries averted it will be biased in the manner we discussed above.

The other possibility is to estimate a totally linear equation

$$s_{it} = AH_{it} + \sum_i B_i 1_i + \sum_t C_t 1_t \quad (\text{A5})$$

This is obviously not the equation we want to estimate, but it is standard and traditional. What bias results if we estimate (A5) instead of (A3)?

Suppose that (A3) is the correct specification, but we estimate (A5) instead. Let h denote the number of neighborhood months in which Homebase is operating, let $t(i)$ denote the number of months that Homebase operated in neighborhood i , and let $i(t)$ denote the number of neighborhoods in which Homebase was operating in month t . Let I denote the total number of neighborhoods, and T the total number of months. If $\hat{a}$ is the OLS estimate of the coefficient of H_{it} in (3), and A is the true coefficient of H_{it} in (A3), then in expected value:

$$\hat{a} - A = - \frac{\sum_i B_i \left\{ \sum_t C_t [It(i) + Ti(t) - h] - IT \sum_{t|(i,t) \in \tilde{H}} C_t \right\}}{h^2 + hIT - I \sum_i t^2(i) - T \sum_t i^2(t)}.$$

(Here the B_i are the true (multiplicative) neighborhood fixed effects in (A3) and the C_t are the month effects.)

The size or direction of the bias is not clear from this expression. For Homebase, however, we can be more specific. Homebase operated in all neighborhoods for some months, and in no neighborhoods for some months. It also operated in some neighborhoods for some months. We can approximate Homebase by the set

$$\tilde{H} = \{(1, 1), (2, 1), \dots, (I, 1), (1, 2), (1, 3), \dots, (1, \tau), \tau < T\}.$$

That is, a program that operated in all neighborhoods in month 1, and then in neighborhood 1 only from month 2 to month τ , and nowhere after month τ . This is simpler than actual Homebase, and the timing is reversed, but the direction of time is immaterial for this calculation, and we are approximating.

Then the bias from estimating (A5) when the true relationship is (A3) is

$$\hat{a} - A = - \left\{ B_1 - \frac{1}{I-1} \sum_{i>1} B_i \right\} \left\{ \frac{C_1 + \sum_{t=\tau+1}^T C_t}{T - \tau + 1} - \frac{\sum_{t=2}^{\tau} C_t}{\tau - 1} \right\}.$$

The first term in curly brackets is the difference between the neighborhood effect in the neighborhood where Homebase operated by itself for a while (approximately the Big Six average) and the average neighborhood effect elsewhere; this is almost certainly positive. The second term in curly brackets is the difference between the average month effect for the period when either all neighborhoods had Homebase or none had it (approximately 2003, early 2004, and 2008), and the average month effect for the period when only the Big Six operated (approximately late 2004 to 2007). This is probably positive (but not assuredly so). Hence $(\hat{a} - A) < 0$. But both $\hat{a}$ and A are likely negative. So Eq. (3) will probably yield an effect of Homebase larger in absolute value (more effective in reducing entries) than Eq. (2)

would, if Eq. (2) were the correct specification. (But with enough pre-program months included in the data set, this bias would diminish, disappear, and eventually reverse.)

One way to reduce the bias from estimating (A5) when (A3) is the true equation is to stratify the sample. The less variation within the data-set in the true counterfactual, the smaller the expected bias from mis-specifying the counter-factual.

A3. Part three—bounds on entries averted

Consider the following model where Homebase has a heterogeneous impact.

$$\ln s_t = X_t - (\beta + \gamma_t)H_t + \varepsilon_t$$

Here s is actual entries, X is the control variables (mainly fixed effects but also foreclosures), β is the average effect of Homebase, H is the dummy variable for Homebase operation (maybe a vector), and γ is the heterogeneous part of the Homebase effect, and ε is the variation in counterfactual entries that would have occurred if HB had not been operating. We just use one subscript for observations rather than two because the distinction between month and CD is not useful here. Let $\mathbf{H}$ denote the set of observations where HB is operating. We assume, as usual,

$$\sum_t \varepsilon_t = 0 \\ \sum_{t \in \mathbf{H}} \gamma_t = 0 = \sum_t \gamma_t H_t \quad .$$

We want to estimate how many entries would have occurred if HB had not been operating.

Let CF_t denote the counterfactual entries in month t : the number that would have occurred if HB had not been operating. Clearly

$$\ln CF_t = X_t + \varepsilon_t$$

But it also happens from the original model that:

$$\ln CF_t = \ln s_t + (\beta + \gamma_t)H_t.$$

Hence

$$CF_t = s_t \exp(\beta + \gamma_t).$$

The total of counterfactual entries is the corresponding sum.

Given β and the observed residuals, what are the bounds on the absolute number of entries averted?

A.3.1 Upper bound

It is easy to see that there is no upper bound. Counterfactual entries are:

$$\sum_{t \in H} C F_t = e^\beta \sum_{t \in H} s_t e^{\gamma_t}$$

The problem of an upper bound is to choose a vector of γ to maximize this expression subject to the constraint that

$$\sum_{t \in H} \gamma_t = 0.$$

Take any two neighborhood-months 1 and 2, with $\gamma_1 = \gamma_2$. Set $\gamma_1 = x > 0$, $\gamma_2 = -x$, $\gamma_t = 0$, $t = 1, 2$. For any x , this vector satisfies the constraint. As x increases, the value of the maximand increases. This continues without any bound.

A.3.2. Lower bound

The objective function and the constraint are the same as above, but second order conditions are satisfied, and so we can set up the Lagrangian:

$$\mathcal{L} = e^\beta \sum_{t \in H} s_t e^{\gamma_t} - \lambda \sum_{t \in H} \gamma_t.$$

Dropping the constant, the derivatives of the Lagrangian are

$$\frac{\partial \mathcal{L}}{\partial \gamma_t} = s_t e^{\gamma_t} - \lambda = 0.$$

Hence for all t , $s_t e^{\gamma_t} = \lambda$. Hence $s_t e^{\gamma_t}$ is a constant. Take logs:

$$\ln s_t + \gamma_t = \ln \lambda.$$

Sum these over H , where n is the number of neighborhood-months in H :

$$n \ln \lambda = \sum \ln s_t + \sum \gamma_t = \sum \ln s_t.$$

Hence

$$\ln \lambda = \frac{1}{n} \sum \ln s_t.$$

So λ is the geometric mean of actual entries when Homebase is operating. The counterfactual for each neighborhood-month is thus

$$CF_t = e^{\beta} \lambda.$$

The counterfactual is just the sum of these values. The number of entries averted is the difference between this sum and actual entries.

References

- Aaronson D. A note on the benefits of homeownership. *J Urban Econ.* 2000; 47(3):356–369.
- Apicello J. A paradigm shift in housing and homeless services: applying the population and high-risk framework to preventing homelessness. *Open Health Serv Policy J.* 2010; 3:41–52.
- Apicello, J.; McAllister, W.; O'Flaherty, B. Homelessness: prevention. In: Smith, S.; Marja-Elsinga; O'Mahony, L.; Eng, O.; Wachter, S., editors. *International Encyclopedia of Housing and Home.* Elsevier; Oxford: 2012.
- Astone N, McLanahan S. Family structure, residential mobility and school dropout: a research note. *Demography.* 1994; 31(4):575–584. [PubMed: 7890092]
- Bassuk E, Buckner J, Weinreb L, et al. Homelessness in female-headed families: childhood and adult risk and protective factors. *Am J Public Health.* 1997; 87:241–248. [PubMed: 9103104]
- Buckner, J. Impact of homelessness on children. In: Levinson, D., editor. *Encyclopedia of Homelessness.* Berkshire Publishing Group; Thousand Oaks, CA: 2004.
- Burt, M.; Pearson, C.; Montgomery, A. *Strategies for Preventing Homelessness.* U.S. Department of Housing and Urban Development; Washington, DC: 2005.
- Cragg M, O'Flaherty B. Do homeless shelter conditions determine shelter population? The case of the Dinkins deluge. *J Urban Econ.* 1999; 46:377–415.
- Culhane DP, Metraux S. Rearranging the deck chairs or reallocating the lifeboats?: homelessness assistance and its alternatives. *J Am Plann Assoc.* 2008; 74(1):111–121.
- Culhane DP, Metraux S, Park JM, Schretzmann M, Valente J. Testing a typology of family homelessness based on patterns of public shelter utilization in four U.S. jurisdictions implications for policy and program planning. *Hous Policy Debate.* 2007; 18(1):1–28.
- Goodman S, Messeri P, O'Flaherty B. How effective homelessness prevention impacts the length of shelter spells. *J Hous Econ.* 2014; 23(1):55–62. [PubMed: 24610995]
- Haveman R, Wolfe B, Spaulding J. Childhood events and circumstances influencing high school completion. *Demography.* 1991; 28(1):133–157. [PubMed: 2015940]
- Johnson, G.; Scutella, R. *Psychological distress and homeless duration*, Working paper, Institute of Applied Economic and Social Research. University of Melbourne; 2014.
- Lazaryan, N.; Ackerman, M.; Neelakantan, U. Does foreclosure increase the likelihood of homelessness? Evidence from the Greater Richmond Area, Working paper. Federal Reserve Bank of Richmond; 2014.
- Main, T. *Homelessness in New York City: Policymaking from Koch to de Blasio.* New York University Press; New York: 2016.
- Mohanty L, Raut L. Home ownership and school outcomes of children: evidence from the PSID child development supplement. *Am J Econ Soc.* 2009; 68(2):465–489.

- New York City Independent Budget Office. Homelessness prevention spending by agency, supplement to Has the rise in homeless prevention spending decreased shelter population? Inside the Budget. 2008 Aug 8. number 157.
- New York City Independent Budget Office. Letter to Patrick Markee. 2012 Jun 14.
- New York City Independent Budget Office. The rising number of homeless families in NYC, 2002-2012: a look at why families were granted shelter, the housing they had lived in, and where they came from. Fiscal brief. 2014 Nov. Available at http://www.ibo.nyc.ny.us/iboreports/2014dhs_families_entering_NYC_homeless_shelters.html
- New York City Independent Budget Office. Albany shifts the burden: as the cost for sheltering the homeless rises, federal and city funds are increasingly tapped. Fiscal Brief. 2015 Oct.
- O'Flaherty, B. Working paper DP0809-14. Columbia University Department of Economics; 2009. What shocks precipitate family homelessness?.
- O'Flaherty B. Individual homelessness: entries, exits, and policy. *J Hous Econ.* 2012; 21(2):77–100.
- O'Flaherty B, Wu T. Fewer subsidized exits and a recession: how New York City's family homeless shelter population became immense. *J Hous Econ.* 2006; 15(2):99–125.
- Remler, D.; Van Ryzin, G. *Research Methods in Practice: Strategies for Description and Causation.* second. Sage; Los Angeles: 2014.
- Rolston, H.; Geyer, J.; Locke, G. Abt Associates. Final Report: Evaluation of the Homebase Community Prevention Program. Department of Homeless Services; New York: 2013.
- Salit S, Kuhn E, Hartz A, Vu J, Mosso A. Hospitalization costs associated with homelessness in New York City. *N Engl J Med.* 1998; 338:1734–1740. [PubMed: 9624194]
- Shinn, M.; Weitzman, B. Homeless families are different. In: Baumohl, J., editor. *Homelessness in America.* Oryx Press; Phoenix: 1996.
- Shinn M, Weitzman B, Stojanovic D, et al. Predictors of homelessness among families in New York City: from shelter request to housing stability. *Am J Public Health.* 1998; 88(10):1651–1657. [PubMed: 9807531]
- Shinn M, Greer A, Bainbridge J, Kwon J, Zuiderveen S. Efficient targeting of homelessness prevention services for families. *Am J Public Health.* 2013; 103(52):S324–S330. [PubMed: 24148041]
- U.S. Department of Housing and Urban Development Office of Community Planning and Development. The 2008 Annual Homeless Assessment Report to Congress. 2009. Available at <https://www.hudexchange.info/resources/documents/4thHomelessAssessmentReport.pdf>
- U.S. Department of Housing and Urban Development Office of Policy Development and Research. Homeless Populations and Sub-population Homeless Populations Report NY-600. 2012a. Available at https://www.hudhre.info/CoC_Reports/2012_ny_600_pops_sub.pdf
- U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Volume I of the 2012 Annual Homeless Assessment Report. 2012b. Available at <https://www.onecpd.info/resource/2753/2012-pit-estimates-of-homelessness-volume-1-2012-ahar/>

Fig. 1. Shelter use by community districts and census tracts, rates of crowding and poverty rate. (Thickened outlines show location of Big Six CDs). *Sources:* New York City Department of Homelssness Services; American Community Survey (2009–2013).

Fig. 2.
Monthly family entries into the New York City shelter system (2003–2008).

Table 1

Annual trends in shelter entries and Homebase cases.

	2003	2004	2005	2006	2007	2008 ^a
	All CDs					
Shelter entries	6481	6726	5951	8116	8190	9647
Shelter entries from CDs during months of official HB operations	<i>b</i>	199 ^c	1396	1915	3692	9647
Total HB cases opened	<i>b</i>	161 ^c	2440	3042	2111	3191
Official HB cases opened		150	2312	2897	1764	3191
(Shelter entries)-to-(official HB cases) ratio ^d		1.33	0.60	0.66	2.09 ^e	3.02
Units starting foreclosure process 2010 population	16,304	13,410	12,886	17,914	27,654	28,132
Total			8,171,680			
Average			138,503			
	"Big Six" CDs					
Shelter entries	1514	1594	1396	1915	1986	2356
Shelter entries from CDs during months of official HB operations		199	1396	1915	1986	2356
Official HB cases opened		150	2312	2897	1404	822
(Shelter entries)-to-(HB cases) ratio 2010 population		1.33 ^f	0.60	0.66	1.41 ^g	2.87
Total			786,814			
Average			131,136			

^a Study period ends November, 2008.^b HB not operating.^c Limited HB operations begin in "Big Six" CDs in November 2004.^d Ratio is restricted to shelter entries and HB cases opened for CDs during official HB operations.^e For first six months of 2007 shelter-HB ratio = 0.86; after expansion of HB program starting in July 2007 Shelter-to-HB ratio = 3.73.^f Ratio is based on 199 shelter entries for November to December period of HB operations.^g For first half of 2007 shelter-HB ratio = 0.86; second half of 2007 shelter-HB-ratio = 2.80.

Table 2

Effects of HB on monthly shelter entries, CD & CT results (point estimates and 95% confidence intervals).

	Loglinear models		Stratified linear models						
	CD-level model	CT-level model	CD-level		CT-level		High-use model	Moderate-use model	High-use model
			Low-use model	High-use model	Low-use model	Moderate-use model			
HB Coverage Variables									
Official operations	-0.052 (-0.15, 0.05)	-0.050 (-0.12, 0.02)	-0.256 (-1.05, 0.65)	-2.040 [†] (-4.26, 0.18)	-0.015 [†] (-0.03, 0.00)	-0.017 (-0.06, 0.02)	-0.013 (-0.17, 0.15)		
Unofficial operations	-0.060 [*] (-0.11, -0.01)	-0.027 (-0.07, 0.02)	-0.257 (-0.70, 0.06)	-0.656 (-2.17, 0.86)	0.007 [*] (0.00, 0.01)	-0.027 [*] (-0.05, -0.00)	-0.056 (-0.17, 0.06)		
Experienced operations	0.000 (-0.10, 0.10)	0.028 (-0.04, 0.10)	-0.236 (-1.05, 0.58)	0.984 (-1.21, 3.18)	0.005 (-0.01, 0.02)	0.009 (-0.04, 0.06)	0.021 (-0.14, 0.18)		
Sum of HB coefficients	-0.112 ^{**} (-0.19, -0.04)	-0.049 [†] (-0.11, 0.01)	-0.749 [†] (-1.51, 0.02)	-1.712 [†] (-3.53, 0.11)	-0.003 (-0.04, 0.03)	-0.035 [*] (-0.07, 0.00)	-0.047 (-0.11, 0.02)		
Sum of foreclosure coef.	0.001 [†] (-0.00, 0.00)	0.016 [*] (0.003, 0.030)	0.038 ^{**} (0.03, 0.05)	0.004 (-0.01, 0.02)	0.013 ^{**} (0.01, 0.02)	-0.000 0.023 ^{**} (0.01, 0.03)	0.01 (-0.02, 0.04)		
Within R ²	0.20		0.20	0.55	0.009	0.030	0.082		
Monthly observations	4189	131,279	2698	1491	70,645	49,832	13,632		
CDs/CTs	59	1849	38	21	995	702	192		

[†] $p < 0.1$.^{*} $p < 0.05$.^{**} $p < 0.01$.

Table 3

Spill over table (CD).

	Loglinear model	Stratified linear model	
		Low use CD	High use CD
Official operations	-0.045 (-0.15, 0.06)	-0.171 (-0.99, 0.64)	-2.080* (-4.29, 0.13)
Unofficial operations	-0.043* (-0.09, 0.01)	-0.174 (-0.52, 0.17)	-0.458 (-1.98, 1.07)
Experienced operations	0.012 (-0.09, 0.11)	-0.135 (-0.95, 0.68)	1.073 (-1.12, 3.26)
Operations next door	-0.075** (-0.12, -0.03)	-0.445* (-0.79, -0.10)	-1.394* (-2.60, -0.19)

[†]
 $p < 0.1$.

*
 $p < 0.05$.

**
 $p < 0.01$.

Table 4

Policy refinement: sum of operating coefficients and unofficial start of operations for loglinear models.

	Use stratum		Season	
	CD	CT	CD	CT
Low use	-0.166** (-0.25, -0.09)	0.130** (0.05, 0.21)		
Moderate use		-0.029 (-0.07, 0.02)		
High use	-0.059 (-0.14, 0.02)	-0.052* (-0.10, -0.01)		
Non-summer			-0.121** (-0.20, -0.04)	-0.033 [†] (-0.075, 0.01)
Summer			-0.080 (-0.20, 0.04)	-0.016 (-0.09, 0.06)
Unofficial	-0.053* (-0.10, -0.00)	-0.037 (-0.08, 0.01)	-0.06* (-0.11, -0.01)	-0.028 (-0.074, 0.018)
	Home base location		Cohort	
	CD	CT	CD	CT
Home base center	-0.058 (-0.14, 0.02)	-0.055 [†] (-0.11, 0.00)		
Adjacent	-0.087 [†] (-0.18, 0.00)	-0.052 (-0.12, 0.02)		
Isolated	-0.186** (-0.27, -0.10)	-0.054 (-0.12, 9.017)		
Big Six pre-2007			-0.044 (-0.14, 0.05)	-0.045 (-0.11, 0.02)
Big Six post-2007			0.028 (-0.10, 0.16)	0.014 (-0.07, 0.10)
2007 cohort			-0.163** (-0.26, -0.07)	-0.073 [†] (-0.15, 0.01)
2008 cohort			-0.046 (-0.16, 0.06)	0.033 (-0.05, 0.12)
Unofficial	-0.047 [†] (-0.10, 0.00)	-0.028 (-0.07, 0.02)	-0.061 (-0.11, -0.01)	-0.027 (-0.07, 0.02)

[†]
 $p < 0.1$.*
 $p < 0.05$.**
 $p < 0.01$.

Table 5

Shelter entries and Homebase cases opened by Homebase center location and cohort.

Operating condition	Shelter entries per CD-month	HB cases per CD-month
Unofficial	11.02	0.54
Has HB center	27.16	23.59
Adjacent	14.57	3.48
Isolated	5.60	1.28
Big Six pre-2007	22.79	34.04
Big Six post-2007	32.89	12.78
2007 cohort	0.32	3.19
2008 cohort	14.70	4.34