Estimating Smoke Alarm Effectiveness and Spatial Distribution in Homes

Dr. Stanley Gilbert

NIST

**National Institute of
Standards and Technology**
U.S. Department of Commerce

NIST Technical Note 2020

Estimating Smoke Alarm Effectiveness and Spatial Distribution in Homes

Dr. Stanley Gilbert
Office of Applied Economics
Engineering Laboratory

November 2018

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Abstract

This report sets out to answer three questions. What effect does installing smoke alarms have on reported fires and casualties for the "average" household? How much of an effect would increasing smoke alarm utilization have on the total number of fires and casualties in the United States? And what are the attributes of homes currently without smoke alarms and where are they located?

We estimate the installation of smoke alarms in homes without them reduces the expected number of fires reported from a (formerly) non-smoke-alarm residence by a factor of 3.5 to 5 and reduces the number of expected casualties by a factor of 2.5 to 3.5.

Unexpectedly, we find the number of casualties per reported fire is lower for non-smoke-alarm households compared to smoke-alarm households. This could be due to changes in people's behavior when they have a smoke alarm in the house or because the less dangerous fires are preferentially extinguished when smoke alarms are present.

If smoke alarms were installed in all residences, the number of fires reported to the fire department could be reduced by 25 % or more. More realistically, each percent increase in smoke alarm penetration reduces reported fires by more than 2.6 %.

Finally, estimates of smoke alarm utilization at the census tract level were developed for the entire country.

We find evidence that the number of homes with smoke alarms are lower than current estimates that rely on telephone-only survey interviews. We estimate actual smoke alarm usage is likely below 92 %. It seems likely that this is due to a correlation between phone presence and smoke alarm utilization.

Key words

Smoke detectors; smoke alarms; small area estimation; home fires

Table of Contents

List of Tables

List of Figures

1. Introduction

Between 2009 and 2013, 2470 people per year lost their lives in home structure fires ('home fires'), and an additional 13 300 were injured, on average [1]. Home fires represented 27 % of all reported fires, yet constituted 84 % and 77 % of all fire fatalities and injuries, respectively. Smoke alarms have been in general use since the early 1970's, and "provide a critical early warning of fire, allowing additional time to escape" [2]. Moreover, "mounting evidence suggests that smoke alarms[1] play a key role in reducing the number of deaths and injuries associated with household fires each year." [3] Some 96 % of homes are reported to have smoke alarms, but more than 35 % of fires occur in homes without smoke alarms. [2] That suggests that installing smoke alarms can have a large effect on the fire problem. It also suggests that the small number of residences without smoke alarms produce a large percentage of the fires.

This report sets out to answer the following questions:

- What effect does installing smoke alarms have on reported fires and casualties for the "average" household?

- How much of an effect would increasing smoke alarm utilization have on the total number of fires and casualties in the United States?

- What are the attributes of homes currently without smoke alarms and where are they located?

In this study smoke alarms are considered to be present without regard to their functional status. Identifying the impact of functional versus non-functional detectors is beyond the scope of this study.

This report differs from previous research in providing better estimates of the impact of installing smoke alarms on reported fires and casualties for the average household. Also, this is the only study that estimates the spatial distribution of smoke alarms in the United States.

The remainder of Sec. 1 contains a literature review and abstract risk model of fire and casualties, taking into account smoke alarms. Sec. 2 estimates the average and total effect of installing smoke alarms on reported fires and casualties. Sec. 3 determines the spatial distribution of smoke alarms. Sec. 4 concludes.

1.1. Literature Review
1.1.1 Smoke Detector Effectiveness

Several studies (e.g., [4] [3] [5]) that evaluated the impact of smoke alarms found that their presence was important for reducing fire risk. A higher percentage of fires are reported to the fire department in homes without smoke alarms than in homes with them. Ahrens [2] reports that "in 2009-2013, more than one-third (35 %) of reported home fires occurred in properties with either no smoke alarms or no working smoke detectors," which constitute less than 5 % of all households. It is worth noting that these studies relate to reported fires. It is not expected that installation of smoke alarms *per se* will reduce ignitions. Rather it studies like

---

[1] Although the literature makes a distinction between smoke alarms and smoke detectors, the terms are used interchangeably in this Note.

Ahrens' suggests that smoke alarms reduce the number of ignitions that get reported to the fire department.

Fires in homes without smoke alarms are deadlier than fires in homes with them. Ahrens reports that "three out of five (59 %) home fire *deaths*[sic] resulted from fires [in homes] without the protection of a working smoke detector." For comparison, only one third of fires occur in homes without working smoke alarms. She also reported that "the death rate per 100 reported fires was twice as high in homes without a working smoke alarm (1.18 deaths per 100 fires) as it was in home fires with this protection (0.53 deaths per 100 fires)."

Runyan et al. [5] studied fatal fires that occurred in North Carolina between January 1988 to January 1989. They compared households where a fatal fire had occurred to a random sample of control households where a non-fatal fire had occurred. They found that "the absence of a smoke detector was a risk factor for fatality, with an odds ratio of 3.4 (95 percent confidence interval, 2.1 to 5.6)." The risk ratio for that death rate was approximately 1.5.

Istre et al. [4] studied the occurrence of fires in relation to smoke alarms in a section of Dallas between 1991 and 1997. They found that 30 % of house fires had smoke alarms present. "The number of injured persons per 100 house fires was higher in house fires without a functioning smoke detector than in those with a functioning smoke detector." Overall (accounting for both the occurrence of fire and the likelihood of injury conditional on occurrence of a fire), "people living in houses without functioning smoke detectors were more than eight times as likely as others to have an injury related to a house fire."

In 2005 the US Consumer Product Safety Commission (CPSC) sponsored a survey to study smoke alarm use in the United States [6]. It had the additional objectives of identifying how many fires occur in the United States and determining what percentage of those fires are reported to the fire department. They found that 96.7 % of households reported having smoke alarms. They estimated that there are 6.6 fires (reported and unreported) per hundred households per year in the United States. Of these only about 3.4 % are reported to the fire department. They found a statistically significant difference in the number of households with smoke alarms between fire households and non-fire households. That implies that households without smoke alarms have more fire ignitions than households with smoke alarms.

The American Housing Survey (AHS) asked about the presence of smoke alarms its 2007, 2009 and 2011 surveys. It found smoke alarm utilization rates ranging from 91.7 % in 2007 to 94.6 % in 2011 (see section 3.1 for more details). These results are different–and lower–than results reported either in the CPSC report or by the NFPA (above). The most likely explanation is response bias. The CPSC and NFPA surveys were telephone surveys. It seems likely that having a phone correlates with having smoke alarms installed in the house. The results below where income and poverty correlate in the expected ways with smoke alarm installation supports that possibility.

1.1.2 Smoke Alarm Utilization

The main data on smoke alarm utilization in the United States comes from Ahrens [2]. She reports the results of a series of surveys between 1977 and 2011, most of which were conducted by the National Fire Protection Association (NFPA), of the presence of smoke

alarms in US households. In 1977, some 22 % of US households surveyed reported having smoke alarms. That increased steadily until 2000 when some 95 % of US households surveyed reported having smoke alarms. From 2001 to 2011, reported smoke alarm presence from surveyed households remained essentially constant between 95 and 97 %. While there is no reason to doubt the trends, evidence suggests that the survey methods used, which relied on telephone interviews, may overestimate the utilization rate for all homes.

Douglas et al. [7] surveyed smoke alarm prevalence in portions of Oklahoma City in 1990 using both a phone survey and an in-person survey. They found that the telephone survey overestimated overall smoke alarm usage when compared to the in-person survey. They concluded that "telephone surveys may overestimate the presence of functioning smoke alarms in some populations" because the populations without telephones are more likely not to have smoke alarms. Further, an analysis of the American Housing Survey data from 2007 – 2011 (described in further detail in Sec. 2 below) produces estimates for smoke alarm utilization significantly below the numbers reported in the smoke alarm surveys.

1.1.3 Increasing Utilization Rates

There have been two studies that evaluated the impact of smoke-alarm give-away programs on fire outcomes. The first study [8] looked at a program in Oklahoma City between 1990 and 1994. They identified a section of the city where fires and fire casualties were high relative to the rest of the city. Their "initial survey indicated that 11 881 of the 34 945 homes in the target area (34 percent) did not have smoke alarms. A total of 10 100 smoke alarms were distributed to 9 291 homes." They found that "the annualized fire-injury rates declined by 80 percent in the target area during the four years after the intervention…, as compared with a small increase in the rest of the city…. There was also a 74 percent decline in the target area in the injury rate per 100 fires (from 5.0 to 1.3; rate ratio, 0.3; 95 percent confidence interval, 0.1 to 0.6), as compared with a small increase in the rest of the city." In the target area the annual rate of fires reported per 1000 homes decreased 25 percent, while in the rest of the city the rate decreased by 18 percent.

The second study [9] looked at a program in Dallas between 2001 and 2011. They installed alarms in homes in at-risk census tracts, with different tracts targeted at different times. The control population was households in the target tracts that were not provided with smoke alarms. They found that the rate of fire casualties was lower for the program population compared to the non-program population (casualty rate per 100 000 population: 3.5 vs 9.5, respectively). "The rates for program versus non-program populations were significantly different for the first 5 years of follow-up … but not for the second 5 years." The effect declined significantly the second five years, most likely due to failure of the smoke alarms.

1.2. Fire Risk Model

We develop a (statistical) fire risk model to evaluate the impact smoke alarms have on deaths and injuries. Fig. 1 shows a conceptual model. The model starts with the probability of ignition and concludes with life-safety outcomes.
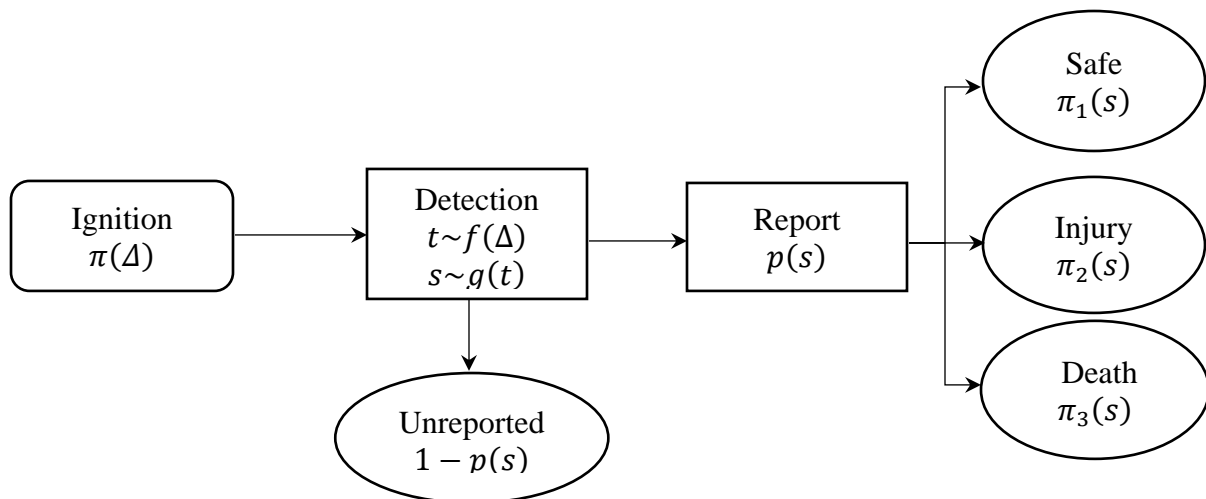
Fig. 1: Model of Detection Time and Home Fires

A natural assumption for modeling ignition is to assume that the probability of ignition (π) is independent of the presence of smoke alarms (where $\Delta \in \{present, absent\}$ indicates whether smoke alarms are present or absent) since smoke alarms are incapable of preventing ignitions. However, the results by the CPSC do not support that assumption. There are at least three mechanisms by which they may be causally linked. First, installation of smoke alarms is a choice by the homeowner / occupant. People who choose to install smoke alarms may also make other choices that impact the probability of fire ignition. Second, the presence of smoke alarms may produce changes in people's behavior that affect the probability of ignition. Third, installation of smoke alarms will correlate with having goods that meet newer flammability standards. In all of these cases it is not the presence of smoke alarms that prevents ignitions; rather it is homeowner / occupant characteristics that correlate with smoke alarms installation.

Detection time (*t*) is a random variable that depends on the presence of smoke alarms. A reasonable assumption is that *holding household factors constant*, installation of a smoke alarm will reduce the detection time. That is, if we exogenously install (or remove) a smoke alarm from a house (taking homeowner/occupant choice out of the equation), the presence of a smoke alarm will on average reduce the detection time.

"Unreported" Fires are assumed to either self-extinguish or be extinguished by the occupants. The vast majority of fires are not reported to the fire department, and (presumably) are extinguished by the occupants. It is expected that the probability that a fire is reported to the fire department ($p_0(s)$) will depend on its size at the time of detection (where *s* is fire size at the time of detection). Fire size at the time of detection will in turn be a random variable that depends on the time to detect. It is assumed for this model that fires that are not reported to the fire department result in no casualties. Otherwise, casualties will depend on the size of the fire at the time of detection.

As mentioned, it is tempting to assume that the presence of smoke alarms is independent of all these steps except for detection. However, we have good reason to believe that this is not true of ignition (likely a result of the endogeneity), so it may not be true of fire size or outcome either.

The empirical models developed and tested in Sec. 2 will be a simplified version of this model.

2. Impact of Smoke Alarms

The purpose of this section is to determine how big an impact the installation of smoke alarms has on the fire problem in the United States. In what follows, Sec. 2.1 describes the data used for this analysis. Sec. 2.2 describes the model analyzed in this section. Sec. 2.3 reanalyzes the CPSC results [7]. Sec. 2.4 describes the results. Sec. 2.5 analyzes an alternate model used for comparison, and Sec. 2.6 concludes with a discussion of the results.

2.1. Data

Data are drawn from the National Fire Incident Reporting System (NFIRS). It is a reporting system used by fire departments nationwide to report on their activities. The system is maintained by the US Department of Homeland Security and is designed to capture all activities engaged in by a fire department, including fires, EMS and community outreach. The system is voluntary at the national level, so some departments do not use the system or report data to it. Data is obtainable from the US Fire Administration.

The NFIRS system records the time, date and location of all incidents, the type of the incident (e.g., fire, EMS call, hazardous materials incident, service call, etc.), property use, equipment and personnel on the call, number type and severity of casualties, actions taken, and a host of other data. For fires specifically, NFIRS collects information on the size of the fire, room of origin, heat source, item first ignited, human and other factors contributing to ignition, presence and effectiveness of alarms and automatic suppression equipment among other data.

Data for NFIRS is filled out by firefighters at the scene, so the information it contains is typically limited to the information a firefighter at the scene would have. For example, fire deaths are defined as any fire casualties resulting in death within one year. However, it is unclear how often follow-up on casualties is revised. Often data that is not required is left unentered. Other systematic problems can occur. For example, a number of large departments report in excess of 80 % of the fires they respond to are confined fires (the nationwide average is less than 40 %). Nevertheless, while NFIRS has known reporting problems, it is still the best data set available for understanding the nature and extent of the fire problem in the United States.

For this analysis we used the 50 departments with the most single-family-residence fires for the years 2009 to 2016. For each incident, the NFIRS data on smoke alarm activation is recorded. Data are categorized based on whether an alarm was present or absent (or unknown). Four departments were excluded *ex ante*: the New York Fire Department (NYFD), Detroit, Baltimore, and Gwinnett County Georgia. The NYFD was excluded because its reporting rate for smoke alarm utilization was below the 30 % limit set below. Detroit was excluded because it was expected based on previous experience to be highly unrepresentative of the rest of the country. Baltimore and Gwinnett County were excluded because they reported no fires where there were no smoke alarms. The analysis below assumes that reporting of smoke alarm status at fires is independent of whether smoke alarms are present, and that assumption was clearly violated for Baltimore and Gwinnett County.

After the top 50 departments were selected, any department with an average reporting rate of smoke alarm usage of less than 30 % across all years of the sample was dropped because it was judged that they would not have enough observations for reliable results. That left 44 departments in the sample. Fig. 2 shows the locations of those departments.
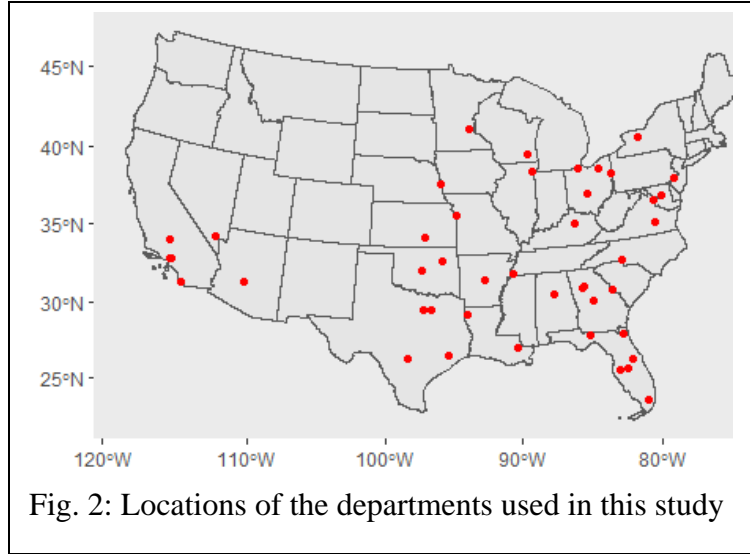
The data selection has the potential to bias the results if the impact of smoke alarms is different for areas served by large departments as compared to smaller departments, or if reporting rate correlates with impact. It was judged that the gains in terms of analytical tractability and reliability outweighed the risk of bias.

Fig. 2: Locations of the departments used in this study

For the 44 departments in this study, for each year between 2009 and 2016, we counted the number of fires reported to NFIRS in each of the following six categories:

- Smoke alarms present, no casualties;
- Smoke alarms absent, no casualties;
- Smoke alarm state unknown, no casualties;
- Smoke alarms present, casualties present;
- Smoke alarms absent, casualties present; and
- Smoke alarm state unknown, casualties present.

2.2. Estimated Model

The model estimated is a simplified version of the conceptual model set out above. It is assumed that the ignition rates for homes with smoke alarms and those without are the numbers identified in CPSC report (analyzed below). Those numbers are assumed to be constant across all 44 departments and all 7 years of the study. The smoke-alarm utilization rate, which is discussed in more detail below, is also assumed to be known and constant across all departments and years.

The model assumes that the reporting rate on smoke-alarm activation for a particular department depends only on whether a casualty occurred. While reporting rate may vary between departments, within a department it only depends on whether a casualty occurred.

There are six independent variables evaluated here determined by alarm state and presence of casualties. Mathematically, the model is:

7

$$
\begin{aligned}
E\big[F_{dt}^{pf}\big] &= pop_{dt}\,\alpha f r^p \delta^p \rho_d^f \\
E\big[F_{dt}^{pn}\big] &= pop_{dt}\,\alpha f r^p (1-\delta^p)\rho_d^n \\
E\big[F_{dt}^{af}\big] &= pop_{dt}\,(1-\alpha)f\lambda r^a \delta^a \rho_d^f \\
E\big[F_{dt}^{an}\big] &= pop_{dt}\,(1-\alpha)f\lambda r^a (1-\delta^a)\rho_d^n \\
E\big[F_{dt}^{uf}\big] &= pop_{dt}\,[\alpha f r^p \delta^p + (1-\alpha)f\lambda r^a \delta^a](1-\rho_d^f) \\
E\big[F_{dt}^{un}\big] &= pop_{dt}\,[\alpha f r^p (1-\delta^p) + (1-\alpha)f\lambda r^a(1-\delta^a)](1-\rho_d^n)
\end{aligned}
\tag{1}
$$

In this, $F_{dt}^{\Delta k}$ is the number of fires reported to fire department d and year t for alarm status Δ and casualty status k. Here Δ is one of 'p' (indicated that smoke alarms were present), 'a' (indicating that smoke alarms were absent), and 'u' (indicating that the alarm state was unknown). The casualty status, k, is one of 'f' (indicating that a casualty occurred during the fire) and 'n' (indicating that no casualty occurred). The variable pop_{dt} is the number of housing units for department d and year t, and α is the proportion of households with smoke alarms.

The variable f is the reported fire production rate for homes with smoke alarms. It is defined as:

$$
f = \mathrm{E}\left\{\frac{\#Fires}{\#Homes}\,\Big|\text{Alarm Present}\right\}.
\tag{2}
$$

The variable λ is the multiple by which the fire production rate for homes without smoke alarms differs from that for homes with smoke alarms. That is:

$$
f\lambda = \mathrm{E}\left\{\frac{\#Fires}{\#Homes}\,\Big|\text{No Alarm Present}\right\}.
\tag{3}
$$

The variable r^p is the "pass-through rate" for homes with smoke alarms: that is, it is the proportion of fires which are reported to the fire department for homes with smoke alarms. The variable r^a is the pass-through rate for homes where smoke alarms are absent.

The variable δ^p is the probability that a fire produces at least one casualty conditional on there being an alarm present. That is:

$$
\delta^p = \mathrm{P}\{Casualties > 0 | \text{Fire \& Alarm Present}\}.
\tag{4}
$$

The variable δ^a is the probability that a fire produces at least one casualty conditional on there being no alarm present:

$$
\delta^a = \delta^p \theta = \mathrm{P}\{Casualties > 0 | \text{Fire \& No Alarm Present}\}.
\tag{5}
$$

The convenience variable θ is defined as ratio of δ^a and δ^p.

The variable ρ_d^f is the probability that the smoke alarm state is reported for department d conditional on a casualty occurring. That is:

$$\rho_d^f = \text{P\{Alarm State Reported| Fire \& Casualty \& Department} = d\}. \quad (6)$$

Finally, ρ_d^n is the probability that the smoke alarm state is reported for department d conditional on no casualty occurring. That is:

$$\rho_d^n = \text{P\{Alarm State Reported| Fire \& No Casualty \& Department} = d\}. \quad (7)$$

Of these, pop_{dt}, the $F_{dt}^{\Delta k}$, α, λ, and f terms are parameters calculated using existing data. The remainder are parameters to be estimated.

Ahrens [2] computes the casualty ratio as:

$$\frac{\dfrac{F^{af}}{F^{an} + F^{af}}}{\dfrac{F^{pf}}{F^{pn} + F^{pf}}} \quad (8)$$

(where department and year notation is suppressed for clarity), which is different from what is used here. Substituting for the variables above, that becomes:

$$\frac{\dfrac{pop \ (1-\alpha)f\lambda r^a \delta^a \rho^f}{pop \ (1-\alpha)f\lambda r^a (1-\delta^a)\rho^n + pop \ (1-\alpha)f\lambda r^a \delta^a \rho^f}}{\dfrac{pop \ \alpha f r^p \delta^p \rho^f}{pop \ \alpha f r^p (1-\delta^p)\rho^n + pop \ \alpha f r^p \delta^p \rho^f}}. \quad (9)$$

Cancelling common terms and rearranging, this becomes:

$$\frac{\delta^a}{\delta^p} \frac{(1-\delta^p)\rho^n + \delta^p \rho^f}{(1-\delta^a)\rho^n + \delta^a \rho^f}. \quad (10)$$

This expression reduces to the form used in this report if and only if $\rho^n = \rho^f$ or $\delta^a = \delta^p$. That is, the two numbers are identical if and only if the casualty rates are the same regardless of alarm presence or the reporting rate for smoke alarms is the same for fires when there are casualties as when there are not. The probabilities that $\rho^n = \rho^f$ or $\delta^a = \delta^p$ are estimated below.

The random variables $F_{dt}^{\Delta k}$ are assumed to be distributed as Poisson variables. The model is estimated in STAN [11].

As discussed above, it seems likely that the proportion of homes with smoke alarms is overestimated when based on a telephone survey. For that reason, this model is estimated with five different numbers for smoke alarm utilization. Input parameters used for those models is listed in Table 3.

2.3.   Reanalysis of the CPSC Results

In 2005, the Consumer Product Safety Commission (CPSC) funded a survey to determine the characteristics of fire households versus non-fire households and to estimate the number of fires occurring annually [7]. They asked people about occurrence of any fires that may have occurred in their household, regardless of whether those fires were reported to the fire department. They also asked about a number of personal and household characteristics.

Table 1: Response Weights			
		Smoke Alarms	
		Yes	No
Fire	Yes	849	67
	No	2,046	68

They found that there were 6.6 fires per hundred households per year in the United States. Of these, only about 3.4 % were attended by the fire department. The fire production rate was estimated using a sophisticated weighting model, accounting for the weights from the stratified survey and the time since the fire occurred (to account for variable recall ability).

Among other characteristics, they asked about the presence of smoke alarms in the house. They estimated that 96.7 % of households reported having at least one smoke alarm in the house. When they compared fire households to non-fire households they found (among other things) that "92.7 percent of fire households and 96.8 percent of non-fire households had at least one smoke alarm," a difference that they reported as being statistically significant (95 % confidence level).

There are two objectives in reanalyzing the CPSC results. First, is to understand how much higher the ignition rate is for non-smoke-alarm households compared to smoke-alarm households. Second, needed are estimates of confidence intervals for the ratio above as well as the percent of smoke alarm households.

To do that, we estimated the number of survey responses in each of the four categories defined by whether the household had a fire (yes/no) and whether it had smoke alarms (yes/no; see Table 1). The survey was designed so that non-fire households were sampled at a rate of 1:40, so the non-fire households were given a weight of 40.

A Bayesian model (STAN [11]) is used to estimate (1) the fire production rate for homes with smoke alarms, (2) the ratio of the fire production rate for homes without smoke alarms to those with smoke alarms, and (3) the percent of homes with smoke alarms. Input data is

Table 2: Reanalysis results for the CPSC report.
Here f is the fire production rate per household, α is the percent of households with smoke alarms, and λ is the ratio of ignitions for non-smoke-alarm households to smoke-alarm households.

	mean	std dev	Confidence Limits 2.5 %	97.5 %	Beta Distribution Parameters alpha	beta	Gamma Distribution Parameters shape	scale
f	0.0656		0.0546	0.0764			126.0	5.2×10^{-4}
α	0.9673	0.0006063	0.9661	0.9685	83,045	2,800		
λ	2.3773	0.2956346	1.8368	2.9934			64.66	3.68×10^{-2}

Table 3: Model Parameters and results

				Pass Through Rate			Casualties (%)		
	α (%)	f	λ	Alarms	No Det.	Ratio	Alarms	No Det.	Ratio
Model1	96.7	0.0645	2.34	0.0103	0.1250	12.09	3.00	2.06	0.686
Model2	94.0	0.0645	2.34	0.0106	0.0677	6.38	3.00	2.06	0.686
Model3	92.0	0.0645	2.34	0.0108	0.0508	4.69	3.00	2.06	0.686
Model4	90.0	0.0645	2.34	0.0111	0.0406	3.67	3.00	2.06	0.686
Model5	87.5	0.0645	2.34	0.0114	0.0325	2.85	3.00	2.06	0.686

the response numbers from Table 1 (modified by the response weight given to non-fire households). The Bayesian model then found best-fit parameters for f, α, and λ.

Table 2 reports the results from this reanalysis. The results for the fire production rate are simply copied from the CPSC report. This analysis was able to reproduce the results from the CPSC report for smoke alarm utilization. In addition, it was determined that the average house without a smoke alarm had an average of 2.4 fire ignitions for every one ignition in a house with smoke alarms.

The table also includes best-fit parameters for probability distributions for the respective parameters. Results for the smoke-alarm utilization results were fit to a beta distribution, and the other two values were fit to gamma distributions. Those parameters were used in the subsequent analysis.

2.4. Results

Results of this analysis are shown in Table 3, except for the department-by-department estimated reporting rates (which are available from the author on request). The 44 departments in the study have widely varying rates at which they reported smoke alarm state at fires. Reporting rates varied from 29.1 % to 87.9 % for fires at which no casualties occurred, with the "average" department having about a 53 % reporting rate. In addition, seven departments (including the NYFD) were excluded from this study for having reporting rates less than 30 %, and the City of Baltimore (excluded for other reasons) has a reporting rate less than 10 %.

Some 34 out of 44 departments have higher reporting rates when there are casualties compared to when there are none. Of these 18 are statistically significantly higher at the

Fig. 3: Pass-Through rate for fires.
Error bars are 95 % confidence limits.
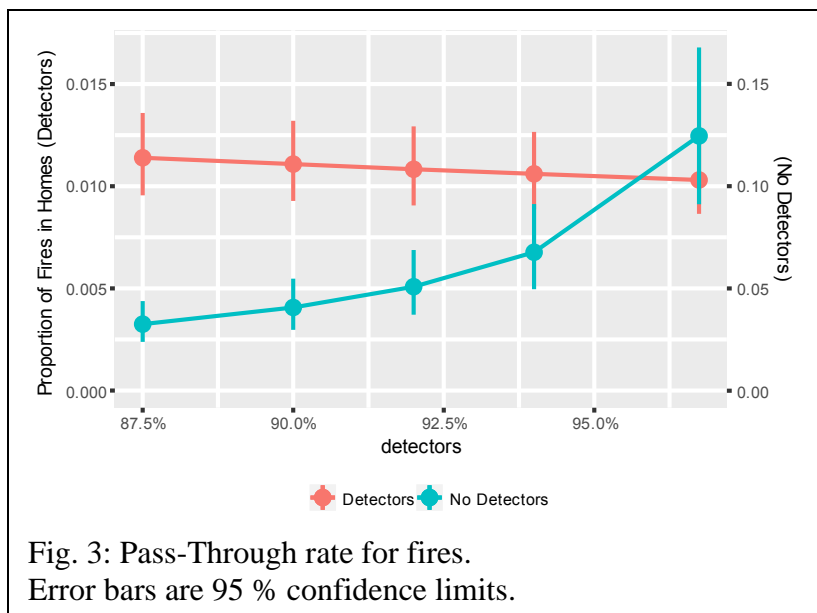
95 % confidence level. None of the 44 departments have statistically significantly lower reporting rates for casualties.

The *r* variable above can be thought of as a "pass-through rate" for fires (i.e., the rate at which ignitions "pass through" and are reported to the fire department). Estimates for its value vary depending on the value attributed to smoke-alarm usage and based on whether smoke alarms are present. As can be seen in Fig. 3, estimates for homes with smoke alarms vary from 0.011 for the case where smoke alarms are in 87.5 % of houses to 0.010 when smoke alarms are in 96.7 % of houses. Estimates of the pass-through rate vary much more dramatically for homes without smoke alarms. Those estimates vary from 0.032 for the case where smoke alarms are in 87.5 % of houses to 0.122 when smoke alarms are in 96.7 % of houses.

Fig. 4 shows the inverse of the pass-through rate (here termed the 'extinguishment rate') and can be interpreted as the number of fires extinguished (by someone other than the fire department or self-extinguish) for every fire that gets reported to the fire department.

Fig. 5 is the ratio of the pass-through rates, with 95 % confidence limits shown. If smoke alarm utilization is 87.5 % then an ignition will be 2.9 times more likely to be reported to the fire department if it occurs in a home without a smoke alarm compared to one with a smoke alarm. That ratio rises to 12.2 when smoke alarm utilization is 96.7 %.

On average, about 3 % of reported fires in houses with smoke alarms produce casualties, a value which does not change with the smoke-alarm utilization rate. Reported fires in
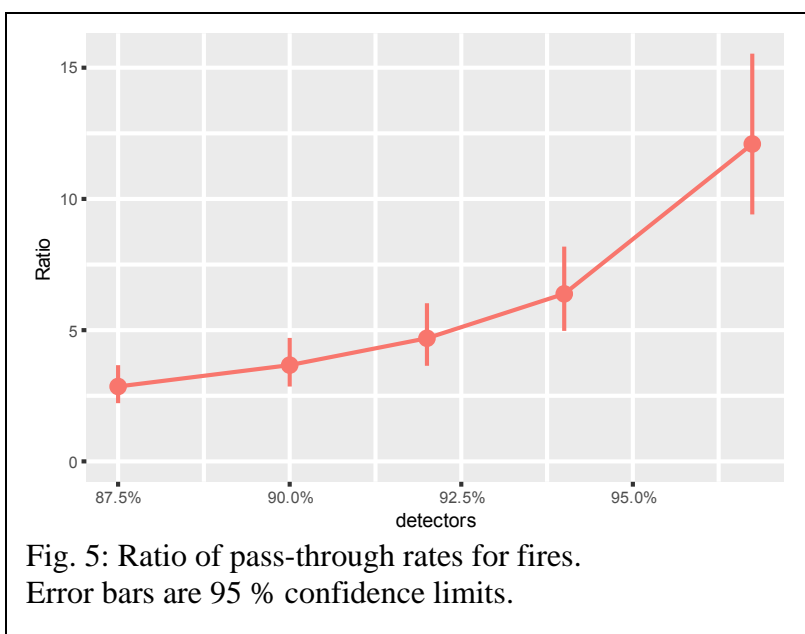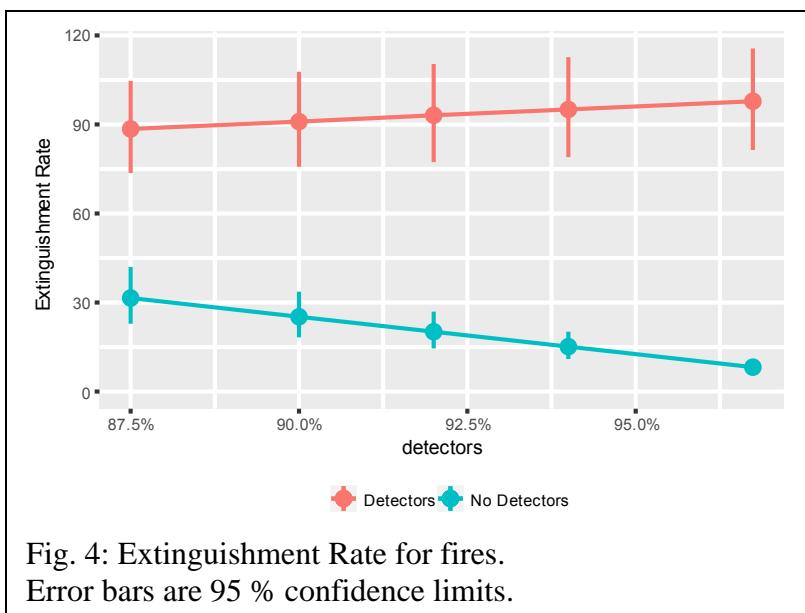
Fig. 4: Extinguishment Rate for fires.
Error bars are 95 % confidence limits.

Fig. 5: Ratio of pass-through rates for fires.
Error bars are 95 % confidence limits.

Table 4: Results of tests

Model	α (%)	$\rho^n = \rho^f$			$\delta^a = \delta^p$			
		W	df	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
Model1	96.7%	607.2	44	$< 10^{-16}$	-0.0094	0.0012	-8.114	4.89×10^{-16}
Model2	94.0%	607.7	44	$< 10^{-16}$	-0.0095	0.0012	-8.125	4.47×10^{-16}
Model3	92.0%	594.1	44	$< 10^{-16}$	-0.0094	0.0012	-8.121	4.63×10^{-16}
Model4	90.0%	616.4	44	$< 10^{-16}$	-0.0094	0.0012	-8.133	4.16×10^{-16}
Model5	87.5%	614.5	44	$< 10^{-16}$	-0.0094	0.0011	-8.246	1.64×10^{-16}

houses without smoke alarms are *less* likely to produce casualties than houses with smoke alarms, by a factor of about 0.7. This differs substantially from the results of Ahrens [2].

As shown in Table 4, reporting rate differs significantly between casualty and non-casualty fires. The hypothesis tested is that the reporting rate for each department is the same regardless of whether there are casualties, that is, for each department d, $\{\rho_d^n = \rho_d^f\}$, and uses the Wald test. Similarly, the results in Table 4 indicate that $\delta^a \neq \delta^p$. Therefore estimating the relative effect of alarms on casualties based on raw NFIRS data (i.e., without correcting for differential reporting rates) will be biased.

2.5.   Analysis of Historic Data

One way of setting an upper bound on the impact of smoke alarms is to assume that all the gains in the fire problem since 1977 (when smoke-alarm data are first available) are from the installation of smoke alarms and to estimate what smoke-alarm impact would produce that level of result. Since it seems likely the installation of smoke alarms is only one of several processes and improvements at work to reduce reported fires and fire casualties, this serves as an upper bound on the impact of smoke alarms.

The model estimated is:

$$F_t = pop_t f[\alpha_t + (1 - \alpha_t)\lambda]$$
$$d_t = pop_t f \delta^d [\alpha_t + (1 - \alpha_t)\lambda\theta] \tag{11}$$

Where pop_t is the number of households in the country at time t, α_t is the smoke-alarm utilization rate at time t, f is the *reported* fire ignition rate for homes with smoke alarms, λ is the ratio of *reported* fire rates, δ^d is the *death* production rate for homes with smoke alarms, and θ is the ratio of *death* production rates (italics emphasize the differences from the previous model). The value for λ in this model will be roughly comparable to the λ in the previous model.

Data on number of single-family residential fires and casualties is from Haynes (2015) [1]. Data on number of households is from the US Census. Two versions of this analysis were run. In the first model, we used the smoke-alarm

Table 5: Results of historical analysis.

	f	λ	δ^d	θ
Model1	0.0037	4.30	0.0091	0.769
Model2	0.0028	5.56	0.0099	0.721

Fig. 6: Modeled fires and deaths: Historical Model 2.
Error bars are 95 % confidence limits.
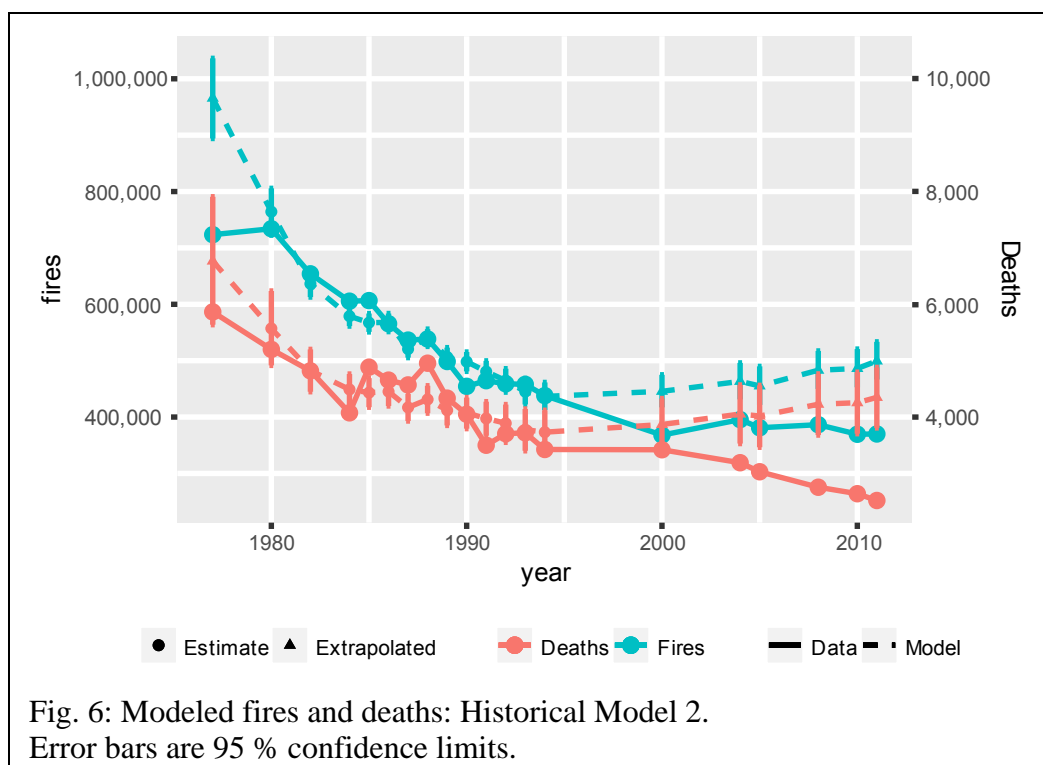
utilization data from Ahrens. In the second model, we rescaled the smoke-alarm utilization data by a factor of 0.9375. That implies that smoke alarm utilization stabilizes after the year 2000 at around 90 % of households. Use of the two models allowed us to bracket the range in which smoke alarm utilization is expected to lie so these results could be compared to the earlier results. In both versions, we restricted our attention to the time-period from 1980 – 1995. That was done because smoke alarm utilization stabilizes by 2000, and any changes in fire rates after that date will be due to other causes. Results are in Table 5.

Fig. 6 shows actual and estimated fires and deaths from 1977 to 2011 for model 2 (based on the rescaled smoke alarm usage estimates). The graph for model 1 is similar.

While the model fits are good for the years 1980 to 1995, they are not very good outside those years. In particular, if smoke-alarm installation were the only contributor to reductions in reported fires and deaths, then fires and deaths would have been increasing since 2000, and they have actually been stable or decreasing. That serves to emphasize that this model provides an upper bound rather than an actual estimate.

It is interesting that the results for ratio of deaths for homes without smoke alarms to homes with smoke alarms for both models are around 0.7, which is similar to the results obtained for casualties in the model above.

The ratio of fires for homes without smoke alarms to homes with smoke alarms ranges for these models between 4.2 (assuming smoke alarm utilization stabilizes at around 96 %) and 5.6 (assuming smoke-alarm utilization stabilizes at around 90 %).

Fig. 7 shows the ratio of pass-through rates estimated from the historical data superimposed on the pass-through rate ratio estimated above. The estimate based on the historical models

14

Fig. 7: Comparison of ratios of Pass-Through Rates.
Error bars are 95 % confidence limits.
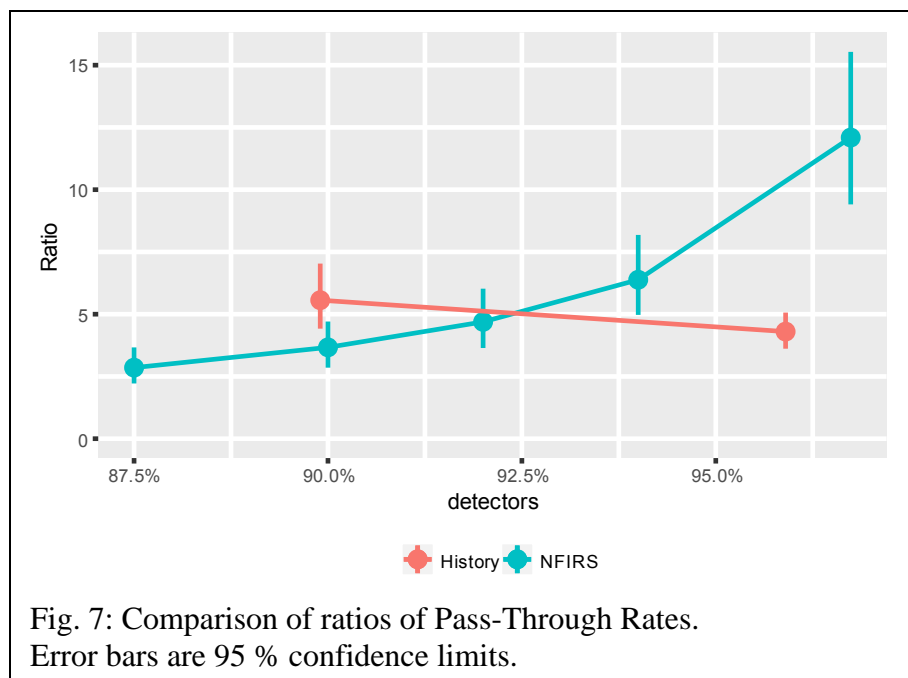
should represent an upper bound on the pass-through ratio for any given value of smoke-alarm utilization, while the "NFIRS" line represents the best estimate of the pass-through ratio for any given value of smoke-alarm utilization. The historical estimate crosses the estimates based on the NFIRS data at about 92.5 %, which suggests that actual smoke-alarm utilization is below 92 %, and the pass-through ratio is below about 5.

2.6. Discussion

The results of Sec. 2.5 were intended to provide an upper bound on impact of smoke alarms on the fire problem, and allows us to derive an upper bound estimate on actual smoke alarm utilization of 92 %. This is lower than the CPSC or NFPA estimates and lower than some of the AHS estimates. There are at least two possible explanations for why previous surveys overestimate smoke alarm utilization. First, the CPSC and NFPA surveys were telephone surveys, and it seems likely that having a phone correlates with having smoke alarms installed in the house. Second is social desirability bias (see [11] for an example). People tend to answer polling questions in ways that (they perceive) make them look good to others. Since there is a social desirability associated with having smoke detectors, there would be a tendency to answer questions about the presence of smoke detectors in the affirmative.

We estimate that smoke alarms reduce the expected number of fires reported by a factor of 3.5 to 5.0 and reduce the number of expected casualties by a factor of 2.5 to 3.5.

The result that the number of casualties per reported fire is lower for non-smoke-alarm households compared to smoke-alarm households is unexpected. There are at least a couple of possible (and non-exclusive) explanations for it. First, people, when they know that they do not have smoke alarms (and remember, these numbers come from surveys), could take actions that reduce the life-risk from fires that they do not take if they have smoke alarms installed. Secondly not all fires are created equally. It is possible that the fires which would

be extinguished if smoke alarms were present (but were not) are on average the less dangerous ones.

If smoke alarms utilization were 100 %, the number of fires reported to the fire department could be reduced by 30 % or more and casualties by 20 % or more. More realistically, each percent increase in smoke alarm utilization reduces reported fires by more than 2.6 % and casualties by more than 1.5 %.

3. Spatial Distribution of Smoke Alarms

This section describes the model used to estimate the spatial distribution of smoke alarm utilization in the United States. It also describes the results of the estimated distribution. In what follows, Sec. 3.1 describes the data used in the analysis, while Sec. 3.2 describes the model used to develop the map. Sec. 3.3 describes the results of the estimation. Sec. 3.4 Concludes this section with a discussion of the results.

3.1. Data

Data for this are from three sources: The American Housing Survey (AHS), the American Community Survey (ACS) summary tables, and the American Community Survey Public Use Microdata Sample (PUMS).

The AHS is sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. It is typically conducted every two years and collects information about the housing stock in the United States. In 2007, 2009 and 2011 the survey included questions about the presence of smoke alarms in homes. The estimated percentage of households with smoke alarms nationwide, based on the American

Table 6: Smoke alarm utilization from the AHS

	pct	se (%)
2007	91.7 %	0.195
2009	92.7 %	0.1952
2011	94.6 %	0.1202

Housing Survey, is shown in Table 6. As discussed above, that these results are different–and lower–than results reported either in the CPSC report [6] or by the NFPA [2] (see above).

The ACS is an annual survey of more than 3.5 million people conducted by the US Census. Its purpose is to help local officials, community leaders, and businesses understand the changes taking place in their communities, and to provide detailed population and housing information about the country. The data from the ACS is released in several forms. Two forms are relevant to this work. The Census releases tables with summaries of the most recent 5-years of results at the census tract level of detail. They also release a sample of the actual responses as Public Use Microdata Samples. The PUMS data are extensively modified to protect the confidentiality of responses. The highest spatial resolution available for the PUMS is the Public Use Microdata Area (PUMA), an area that contains between 100 000 and 200 000 people. Each year's PUMS data set contains about 1 % of the US population.

Variables used in the model are listed in Table 7.

Spatial areas in the AHS are not the same as those used in the ACS. The AHS uses the 1983 Metropolitan Statistical Areas (MSA) as the basis for its spatial identifications, and the ACS does not use them at all. The 1983 MSAs are made up of counties (or in the case of the New England states, of county subdivisions) as they existed in 1983. As part of this analysis it was necessary to identify PUMAs and Census Tracts with (1983) MSAs. That was done using a

Table 7: List of variables.
* Excluded to prevent collinearity.

Group	Value	Description	Percentage	Note
tenure	Own	The home is occupied by its owners	68.31 %	*
	Rent	The home is occupied by renters	31.69 %	
units	Apt	Apartment	22.93 %	*
	MH	Mobile Home	6.32 %	
	SFR	Single-Family Residence	70.76 %	
built	1939-	Built before 1940	15.13 %	*
	1940s	Built between 1940 and 1949	5.94 %	
	1950s	Built between 1950 and 1959	10.32 %	
	1960s	Built between 1960 and 1969	11.83 %	
	1970s	Built between 1970 and 1979	18.96 %	
	1980s	Built between 1980 and 1989	12.64 %	
	1990s	Built between 1990 and 1999	12.64 %	
	2000s	Built between 2000 and 2009	12.37 %	
	2010s	Built after 2009	0.17 %	
bedrms	1	Home has one bedroom	10.17 %	*
	0	Home has no bedrooms	0.74 %	
	2	Home has 2 bedrooms	24.81 %	
	3	Home has 3 bedrooms	43.07 %	
	4	Home has 4 bedrooms	16.94 %	
	5+	Home has 5 or more bedrooms	4.28 %	
persons	Other	Occupied by more than 1 person but not crowded	67.27 %	*
	Crowded	More people than rooms	6.09 %	
	Single	Occupied by only one person	26.65 %	
childs	No	No children in home	88.57 %	*
	Yes	Children in home	11.43 %	
poverty	Not	Household income above poverty line	87.16 %	*
	pov1.0	Household income below poverty line	12.84 %	

Variable	Description	Mean
rooms	Number of rooms in home	5.7007
zinc2	Household income	66,099

table from census listing what county (and, for the New England states, county subdivisions) were associated with each MSA. In a few cases, the FIPS codes for the counties or county subdivisions had changed between 1983 and 2013. Those were identified by hand and updated. The 2013 Tiger files for counties and (where relevant) county subdivisions were then unioned to form the MSA area. Each PUMA and tract were associated with the SMSA whose boundary its centroid fell within. Note that some county boundaries have changed between 1983 and the present. This process assumes they have not. Since such changes are likely to be few and small, they are ignored for this work.

3.2.   Analysis
Census-tract-level estimates of smoke-alarm utilization are developed using the following approach.

First, an individual-level (household) logit model for smoke alarms depending on household characteristics is developed from the AHS data.

Then, using the model and matching data from the ACS, the expected number of households with smoke alarms is estimated for census tracts.

That is, we estimate:

$$y_i = \frac{1}{N_i} \sum_{j \in I_i} f(x_j) \tag{12}$$

for each tract, i, where I_i is the set of households in tract i, N_i is the number of households in tract i, and f is the probability that a household has a smoke alarm conditional on x and is the model estimated from the AHS data.

A difficulty arises here: the AHS model is estimated at the individual household level, and the public ACS tables do not provide information at that level, at least not at Census tract detail. Rather the ACS tables effectively provide totals or averages for each census tract. If this were a linear model, there would be no problem, since the average of the model estimates would be equal to the model estimate of the average. However, the underlying logit model is non-linear. Worse, the estimates will be very close to one, which means that the model is concave and has a relatively high curvature. As a result, the estimates will be biased toward 1.

The approach used here is to linearize the model using a second-order Taylor approximation. This provides some non-linearity to a model where nonlinearity is expected to be important while still keeping the model manageable. Then, the expected number of households in a Census tract with smoke alarms is estimated to be:

$$y_i \approx \overline{p}_i + \frac{1}{2} \overline{p}_i (1 - \overline{p}_i)(1 - 2\overline{p}_i) \beta' \Xi_i \beta, \tag{13}$$

Where $\overline{p}_i$ is the value determined from the model estimated above using the average values for tract i, and Ξ_i is the covariance matrix for the covariates used in the estimation for tract i (mathematical details are in Appendix A).

The value for Ξ_i cannot be computed directly for each census tract. However, Census makes a sample of individual-level and household-level information available for PUMAs. For this study, the value for Ξ_i is computed for each PUMA. The value of Ξ_i for a specific Census tract is assumed to be identical to that for the PUMA it nests within.

Table 8: Results of the AHS-Based Smoke Alarm Logit Model

Coefficient	Estimate	Std. Error	Marginal Effect	t value	Pr(>|t|)	Flag[†]
tenureRent	-0.10	0.04	-0.10	-2.85	3.56 %	*
unitsMH	-1.13	0.06	-1.13	-18.28	0.00 %	***
unitsSFR	-0.87	0.05	-0.87	-18.87	0.00 %	***
built1940s	0.08	0.05	0.08	1.69	15.19 %	
built1950s	0.23	0.05	0.23	4.80	0.49 %	**
built1960s	0.36	0.04	0.36	8.17	0.04 %	***
built1970s	0.56	0.04	0.56	12.94	0.00 %	***
built1980s	1.04	0.06	1.04	18.10	0.00 %	***
built1990s	1.52	0.08	1.52	19.54	0.00 %	***
built2000s	1.90	0.07	1.90	26.86	0.00 %	***
built2010s	1.69	0.52	1.69	3.26	2.24 %	*
rooms	0.25	0.02	0.42	13.49	0.00 %	***
bedrms0	-0.15	0.14	-0.15	-1.04	34.52 %	
bedrms2	-0.06	0.05	-0.06	-1.02	35.54 %	
bedrms3	-0.02	0.07	-0.02	-0.23	82.48 %	
bedrms4	-0.16	0.10	-0.16	-1.57	17.73 %	
bedrms5+	-0.60	0.13	-0.60	-4.62	0.57 %	**
personsCrowded	-0.48	0.05	-0.48	-9.72	0.02 %	***
personsSingle	-0.04	0.03	-0.04	-1.23	27.43 %	
childsYes	0.26	0.05	0.26	5.54	0.26 %	**
povertypov1.0	-0.26	0.04	-0.26	-7.20	0.08 %	***
zinc2	4.73E-06	4.69E-07	0.33	10.09	0.02 %	***

SMSA Results are listed in Appendix B.

[†] Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3.3.   Results

3.3.1. AHS Model

Results of the individual-level model are shown in Table 8. "Marginal effect" is the parameter value times the standard deviation of the variable for continuous variables and the parameter value for dummy variables. It provides an indication of the typical magnitude of the effect for a variable.

Being a renter reduces the likelihood that there is a smoke-alarm reported in the house. Mobile homes and single-family residences are less likely than apartment complexes/condos to report smoke alarms, with mobile homes being the least likely of the three to report smoke alarms.

The newer the home the more likely smoke alarms were reported. The more rooms a home has the more likely it is to have smoke alarms reported. The number of bedrooms is mostly not significant to the likelihood of having smoke alarms reported in the home. Note that this is controlling for number of rooms in the house, which will strongly correlate with number of bedrooms.

Residences that are "crowded"—that is, that have more people living in them than rooms—are less likely to have smoke alarms reported. Having only a single person living in the home is not related to the likelihood of reported smoke alarms. Children under 18 living in the house makes it more likely that there will be smoke alarms reported in the house.

Fig. 8: Kernel density of smoke alarm estimates

Income is positively associated with the reporting of smoke alarms while having household income below the poverty line is negatively associated with having smoke alarms reported.

3.3.2. Tract-Level Model

Fig. 8 is a plot of the Kernel Density of smoke alarm utilization for the model-based estimates for census tracts.

The bulk of census tracts in the country are estimated to have smoke alarm utilization between 75 % and 100 %.

Maps showing smoke alarm utilization by census tract for Houston (Fig. 9), Portland, Oregon (Fig. 10), and Kansas City, Missouri (Fig. 11) are shown below.

3.4. Discussion

It would be useful to fire departments to know what portions of their service area have smoke alarms and which do not. Knowing that would help them know which areas are at greater risk for fire and help them target smoke-alarm programs. This work identifies the level of smoke-alarm utilization at the census tract level nationwide, thus fulfilling that objective.
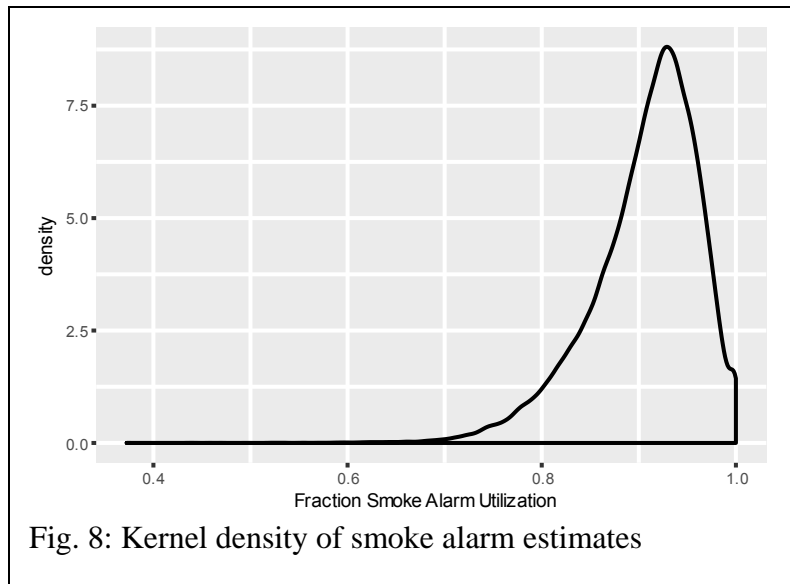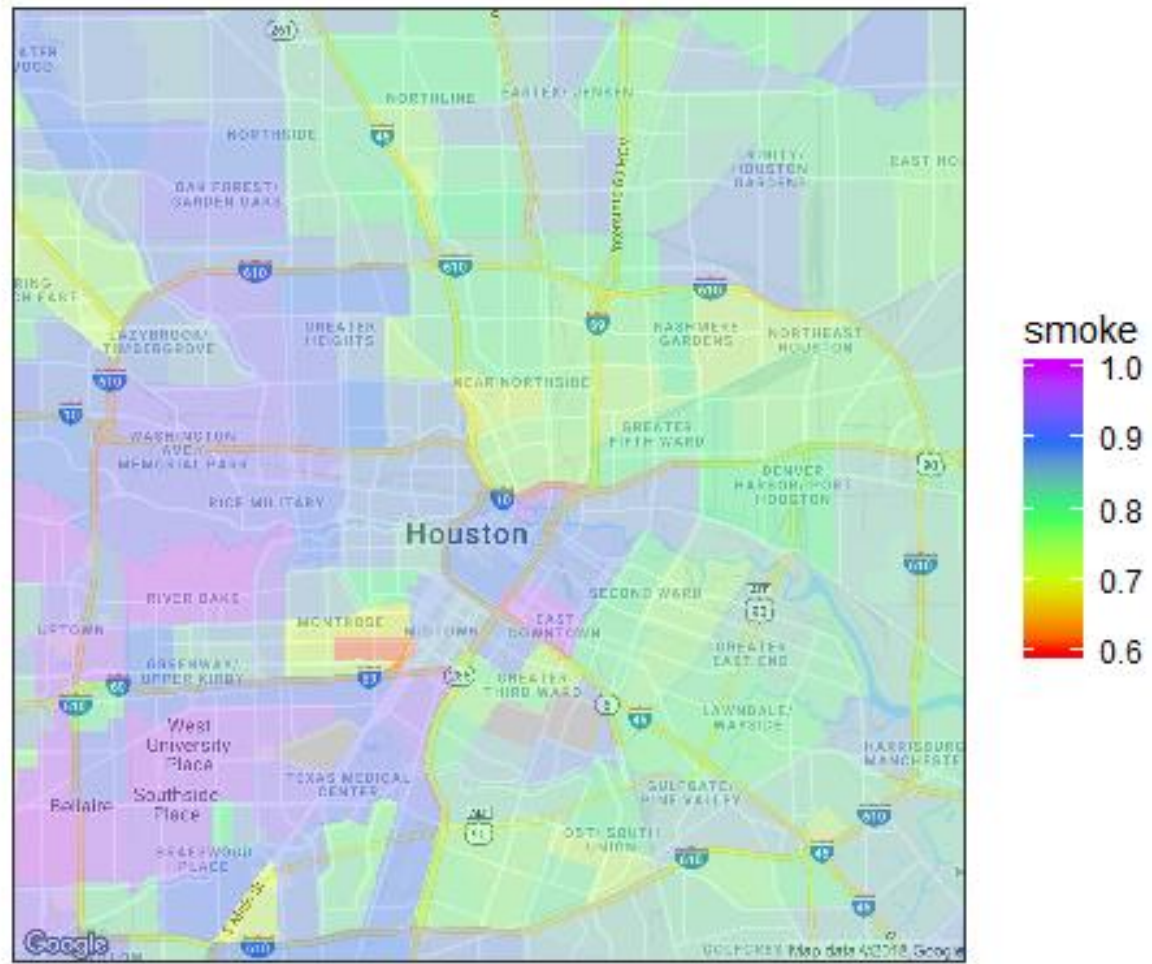
Fig. 9: Smoke alarm utilization by census tract for Houston

Fig. 10: Smoke alarm utilization by census tract for Portland, Oregon.
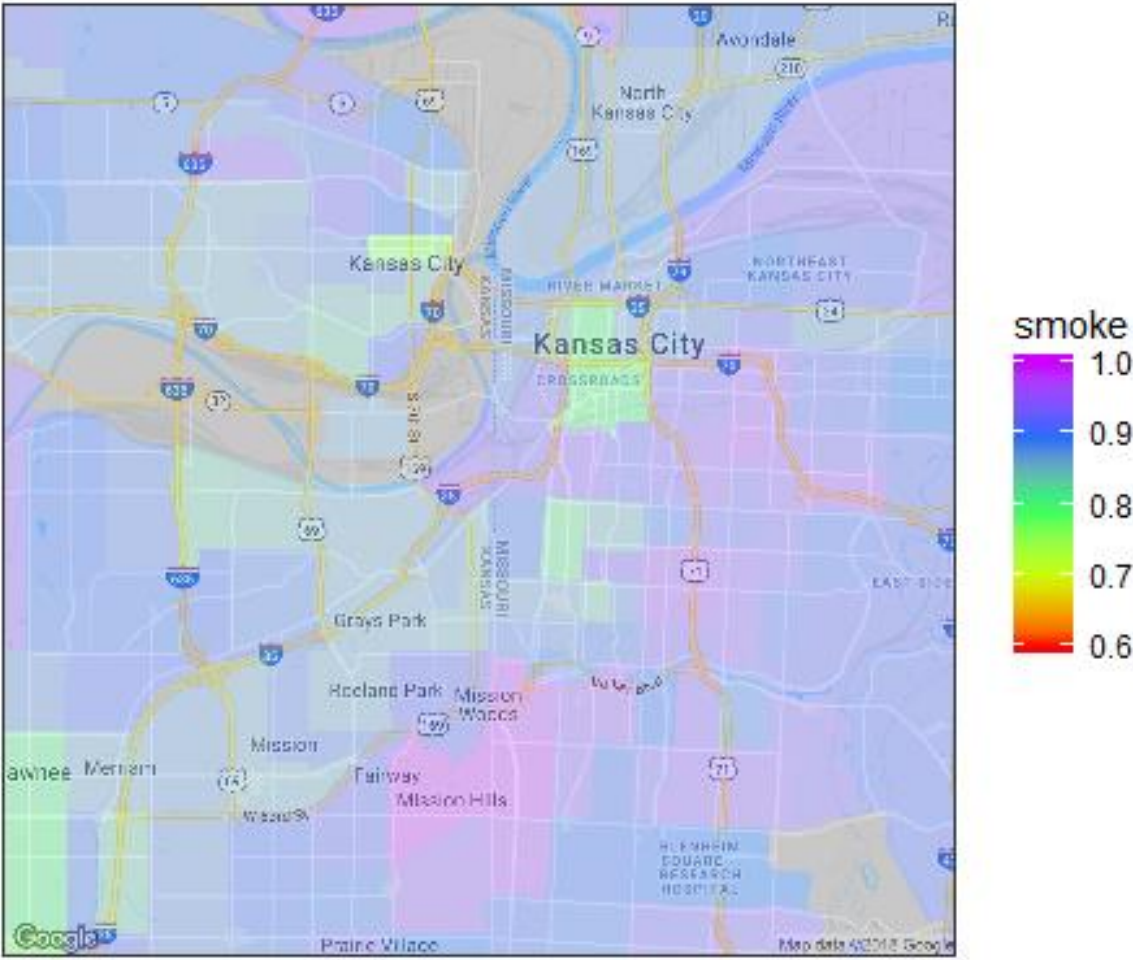
Fig. 11: Smoke alarm utilization by census tract for Kansas City, Missouri

4. Conclusions

This report sets out to find out how big an impact installing smoke alarms has on the fire problem in the United States, how much of an effect installing smoke alarms in the homes without them have on the fire problem, and to determine where homes without smoke alarms are located.

Previous estimates of the number of homes with smoke alarms based on telephone surveys appear biased upward. This is most likely in part because smoke alarm installation likely correlates with having a phone. Actual smoke alarm usage is likely below 92 % Installation of smoke alarms likely reduces the expected number of fires reported from a (formerly) non-smoke-alarm residence by a factor of 3.5 to 5.

The number of casualties per reported fire is lower for non-smoke-alarm households compared to smoke-alarm households, which was an unexpected result. There are a couple of possible (and non-exclusive) explanations for it. First, people, when they know that they do not have smoke alarms (and remember, these numbers come from surveys), they could take actions that reduce the life-risk from fires that they do not take if they have smoke alarms installed. Secondly not all fires are created equally. It is possible that the fires which would be extinguished if smoke alarms were present (but were not) are on average the less dangerous ones. Still, installation of smoke alarms likely reduces the number of expected casualties by a factor of 2.5 to 3.5.

If smoke alarm utilization were 100 %, the number of fires reported to the fire department could be reduced by 25 % or more. More realistically, each percent increase in smoke alarm utilization reduces reported fires by more than 2.6 % % and casualties by more than 1.5 %.

Finally, we develop a mechanism to estimate of smoke alarm utilization at the census tract level for the entire country.

5. References

[1]   H. Haynes, Fire loss in the United States During 2015, National Fire Protection Association, Fire Analysis and Research Division, 2015.

[2]   M. Ahrens, Smoke Alarms in U. S. Home Fires, Quincy, Mass: NFPA, 2015.

[3]   D. Rohde, J. Corcoran and P. Chhetri, "Spatial forecasting of residential urban fires: A Bayesian approach," *Computers, Environment and Urban Systems,* vol. 34, no. 1, pp. 58-69, 2010.

[4]   G. Istre, M. McCoy, L. Osborn, J. Barnard and A. Bolton, "Deaths and Injuries from House Fires," *New England Journal of Medicine,* vol. 344, no. 25, pp. 1911-1916, 2001.

[5]   C. Runyan, S. Bangdiwala, M. Linzer, J. Sacks and J. Butts, "Risk factors for fatal residential fires," *New England Journal of Medicine,* vol. 327, no. 12, pp. 859-863, 1992.

[6]   M. Greene and C. Andres, 2004-2005 National Sample Survey of Unreported Residential Fires, Rockville, MD: US Consumer Product Safety Commission, 2009.

[7]   M. R. Douglas, S. Mallonee and G. Istre, "Estimating the proportion of homes with functioning smoke alarms: a comparison of telephone survey and household survey results.," *American Journal of Public Health,* vol. 89, no. 7, pp. 1112-1114, 1999.

[8]   S. Mallonee, G. Istre, M. Rosenberg, M. Reddish-Douglas, F. Jordan, P. Silverstein and W. Tunell, "Surveillance and prevention of residential-fire injuries," *New England Journal of Medicine,* vol. 335, no. 1, pp. 27-31, 1996.

[9]   G. Istre, M. McCoy, B. Moore, C. Roper, S. Stephens-Stidham, J. Barnard, D. Carlin, M. Stowe and R. Anderson, "Preventing deaths and injuries from house fires: an outcome evaluation of a community-based smoke alarm installation programme," *Injury Prevention,* vol. 20, no. 2, pp. 97-102, 2014.

[10] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li and A. Riddell, "Stan: A probabilistic programming language," *Journal of Statistical Software,* vol. 76, no. 1, 2017.

[11] C. K. Hadaway, P. L. Marler and M. Chaves, "What the polls don't show: A closer look at U.S. church attendance," *American Sociological Review,* vol. 58, no. 6, pp. 741-752, 1993.

Appendix A: Mathematical Model

To estimate tract-level estimates of smoke-alarm usage we start with a model, $f(x)$, estimating the likelihood that an individual household has smoke-alarms given household-specific information, x.

Let $\mathcal{J}$ represent the set of people, and $\{I_i\}_{i=1}^N$ be a partition of the people, where each $i \in \{1 \ldots N\}$ represents a census tract, and N_i represents the total number of households in tract i. What we want to estimate, then, is:

$$Y_i = \sum_{j \in I_i} f(x_j) \tag{14}$$

or

$$y_i = \frac{1}{N_i} \sum_{j \in I_i} f(x_j) \tag{15}$$

for each tract, I, where f is the probability that a household has a smoke alarm, conditional on x.

The problem is that the tract-level information available is summary information and not household level. If the estimated function were linear, this would not be a problem, since, in that case, $\frac{1}{N_i} \sum_{j \in I_i} f(x_j) = f(\bar{x})$ (where $\bar{x}$ is the average value of the information used for estimation). However, the logistic function is non-linear, which means that this approach will be a poor estimate at best and will in fact be biased.

The approach used here to estimating tract-level is to linearize the model using a second-order Taylor approximation. When $y = f(x)$, that takes the form of:

$$y \approx f(x_0) + (x - x_0)' \frac{df(x_0)}{dx} + (x - x_0)' \frac{d^2 f(x_0)}{dx dx'} (x - x_0) \tag{16}$$

This provides some non-linearity to a model where nonlinearity is expected to be important while still keeping the model manageable.

The individual-level estimate of the probability that a household has a smoke-alarm, f, is determined by the logistic function, which will typically be written as:

$$f(\mathrm{x}) = p(x|\beta) = Pr(x|\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}, \tag{17}$$

Where β are the regression parameters estimated in Sec. 3.3.1. For the sake of parsimony, that will typically be written as $p(x)$, suppressing the β, or for some x_j, as $p_j \equiv p(x_j|\beta)$.

Then the value of Y_i for census tract i is:

$$Y_i = \sum_{j \in I_i} f(x_j) \approx \Sigma \left[p_0 + (x - x_0)' \frac{dp(x_0)}{dx} + \frac{1}{2}(x - x_0)' \frac{dp(x_0)}{dx dx'} (x - x_0) \right] \tag{18}$$

Where

$$\frac{dp(x_j)}{dx} = \frac{e^{x_j'\beta}}{\left(1 + e^{x_j'\beta}\right)^2}\beta = p_j(1 - p_j)\beta \tag{19}$$

and

$$\frac{d^2p(x_j)}{dxdx'} = p_j(1 - p_j)(1 - 2p_j)\beta\beta' \tag{20}$$

The revised Taylor series becomes:

$$Y_i = \sum_{j \in I_i} f(x_j) \approx \Sigma \left[p_0 + p_0(1 - p_0)(x_j - x_0)'\beta + \frac{1}{2}p_0(1 - p_0)(1 - 2p_0)(x_j - x_0)'\beta\beta'(x_j - x_0) \right] \tag{21}$$

Or

$$\begin{aligned}
Y_j &\approx N_j p_0 + p_0(1 - p_0)\beta'\Sigma(x_j - x_0) + \frac{1}{2}p_0(1 - p_0)(1 - 2p_0)\Sigma(x_j - x_0)'\beta\beta'(x_j - x_0) \\
&= N_j p_0 + p_0(1 - p_0)\beta'\Sigma(x_j - x_0) + \frac{1}{2}p_0(1 - p_0)(1 - 2p_0)\beta'\left(\Sigma(x_j - x_0)(x_j - x_0)'\right)\beta \\
&= N_j\left(p_0 + p_0(1 - p_0)\beta'\frac{1}{N_j}\Sigma(x_j - x_0) + \frac{1}{2}p_0(1 - p_0)(1 - 2p_0)\beta'\left(\frac{1}{N_j}\Sigma(x_j - x_0)(x_j - x_0)'\right)\beta \right)
\end{aligned} \tag{22}$$

If we let

$$x_0 = \overline{x}_i = \frac{1}{N_i}\sum_{j \in I_i} x_j \tag{23}$$

and note that

$$\Xi_i \equiv \mathrm{Cov}_{j \in I_i}(x) = \frac{1}{N_i}\sum_{j \in I_i}(x_j - \overline{x}_i)(x_j - \overline{x}_i)' \tag{24}$$

and defining $\overline{p}_i = p(\overline{x}_i|\beta)$, this becomes:

$$\begin{aligned}
Y_i &\approx N_i\left(\overline{p}_i + \overline{p}_i(1 - \overline{p}_i)\beta'\frac{1}{N_i}\Sigma(x_j - \overline{x}_i) + \frac{1}{2}\overline{p}_i(1 - \overline{p}_i)(1 - 2\overline{p}_i)\beta'\left(\frac{1}{N_i}\Sigma(x_j - \overline{x}_i)(x_j - \overline{x}_i)'\right)\beta \right) \\
&= N_i\left(\overline{p}_i + \overline{p}_i(1 - \overline{p}_i)\beta'0 + \frac{1}{2}\overline{p}_i(1 - \overline{p}_i)(1 - 2\overline{p}_i)\beta'\Xi_i\beta \right) \\
&= N_i\left(\overline{p}_i + \frac{1}{2}\overline{p}_i(1 - \overline{p}_i)(1 - 2\overline{p}_i)\beta'\Xi_i\beta \right)
\end{aligned} \tag{25}$$

The $\overline{x}_i$ values are available from the ACS summary tables for each census tract. The covariance of x_i is not readily available from the ACS Tables. To obtain that value we use the ACS PUMS data. The 5-year data is resolved to a PUMA (Public-Use Microdata Area) that is significantly smaller than the metropolitan area, but is larger than the Census Tract. Assuming that the variance for the PUMA is the same as that for each Census Tract it contains, we can calculate an approximate Y_i. What is estimated in Sec. 3.3.2, then, is

$$y_i = \frac{Y_i}{N_i} \approx \overline{p}_i + \frac{1}{2}\overline{p}_i(1 - \overline{p}_i)(1 - 2\overline{p}_i)\beta'\Xi_i\beta. \tag{26}$$

Appendix B: AHS Model

Coefficients	Estimate	Std. Error	Marginal Effect	t value	Pr(>|t|)	Flag[†]
tenureRent	-0.10	0.04	-0.10	-2.85	3.56 %	*
unitsMH	-1.13	0.06	-1.13	-18.28	0.00 %	***
unitsSFR	-0.87	0.05	-0.87	-18.87	0.00 %	***
built1940s	0.08	0.05	0.08	1.69	15.19 %	
built1950s	0.23	0.05	0.23	4.80	0.49 %	**
built1960s	0.36	0.04	0.36	8.17	0.04 %	***
built1970s	0.56	0.04	0.56	12.94	0.00 %	***
built1980s	1.04	0.06	1.04	18.10	0.00 %	***
built1990s	1.52	0.08	1.52	19.54	0.00 %	***
built2000s	1.90	0.07	1.90	26.86	0.00 %	***
built2010s	1.69	0.52	1.69	3.26	2.24 %	*
rooms	0.25	0.02	0.42	13.49	0.00 %	***
bedrms0	-0.15	0.14	-0.15	-1.04	34.52 %	
bedrms2	-0.06	0.05	-0.06	-1.02	35.54 %	
bedrms3	-0.02	0.07	-0.02	-0.23	82.48 %	
bedrms4	-0.16	0.10	-0.16	-1.57	17.73 %	
bedrms5+	-0.60	0.13	-0.60	-4.62	0.57 %	**
personsCrowded	-0.48	0.05	-0.48	-9.72	0.02 %	***
personsSingle	-0.04	0.03	-0.04	-1.23	27.43 %	
childsYes	0.26	0.05	0.26	5.54	0.26 %	**
povertypov1.0	-0.26	0.04	-0.26	-7.20	0.08 %	***
zinc2	4.73E-06	4.69E-07	0.33	10.09	0.02 %	***
smsa0080	1.81	0.85	1.81	2.11	8.81 %	.
smsa0160	12.81	0.08	12.81	151.53	0.00 %	***
smsa0200	4.75	0.93	4.75	5.09	0.38 %	**
smsa0240	12.92	0.08	12.92	163.44	0.00 %	***
smsa0360	1.66	0.18	1.66	9.42	0.02 %	***
smsa0460	12.68	0.08	12.68	153.85	0.00 %	***
smsa0520	0.98	0.16	0.98	6.10	0.17 %	**
smsa0560	2.15	1.28	2.15	1.68	15.36 %	
smsa0600	12.80	0.09	12.80	141.59	0.00 %	***
smsa0620	3.22	1.33	3.22	2.43	5.94 %	.
smsa0640	2.35	0.88	2.35	2.69	4.34 %	*
smsa0680	1.72	0.58	1.72	2.96	3.17 %	*
smsa0720	2.56	0.45	2.56	5.73	0.23 %	**
smsa0760	1.76	0.83	1.76	2.11	8.88 %	.
smsa0840	0.18	0.64	0.18	0.28	78.95 %	
smsa0875	2.20	0.54	2.20	4.11	0.93 %	**
smsa1000	1.55	0.27	1.55	5.65	0.24 %	**

Coefficients	Estimate	Std. Error	Marginal Effect	t value	Pr(>|t|)	Flag[†]
smsa1120	3.23	0.33	3.23	9.73	0.02 %	***
smsa1125	1.87	0.60	1.87	3.12	2.64 %	*
smsa1160	1.85	0.42	1.85	4.42	0.69 %	**
smsa1280	2.73	0.18	2.73	14.95	0.00 %	***
smsa1320	3.87	1.26	3.87	3.07	2.78 %	*
smsa1440	1.32	0.43	1.32	3.04	2.88 %	*
smsa1520	1.97	0.15	1.97	12.87	0.01 %	***
smsa1560	1.32	0.29	1.32	4.63	0.57 %	**
smsa1600	2.16	0.13	2.16	16.63	0.00 %	***
smsa1640	2.37	0.29	2.37	8.19	0.04 %	***
smsa1680	2.26	0.19	2.26	11.89	0.01 %	***
smsa1720	1.85	0.45	1.85	4.09	0.94 %	**
smsa1760	1.24	0.50	1.24	2.49	5.52 %	.
smsa1840	2.32	0.28	2.32	8.44	0.04 %	***
smsa1880	0.08	0.35	0.08	0.23	82.37 %	
smsa1920	1.29	0.14	1.29	9.14	0.03 %	***
smsa1960	1.63	0.47	1.63	3.47	1.78 %	*
smsa2020	0.69	0.97	0.69	0.72	50.59 %	
smsa2080	1.22	0.17	1.22	7.01	0.09 %	***
smsa2120	1.47	0.51	1.47	2.91	3.35 %	*
smsa2160	1.98	0.13	1.98	15.64	0.00 %	***
smsa2320	0.80	0.25	0.80	3.23	2.32 %	*
smsa2400	2.90	0.61	2.90	4.76	0.51 %	**
smsa2640	2.07	0.50	2.07	4.13	0.91 %	**
smsa2680	0.85	0.16	0.85	5.20	0.35 %	**
smsa2760	3.11	1.31	3.11	2.37	6.38 %	.
smsa2800	1.86	0.23	1.86	7.96	0.05 %	***
smsa2840	2.14	0.44	2.14	4.89	0.45 %	**
smsa2960	2.18	0.44	2.18	4.98	0.42 %	**
smsa3000	3.02	0.66	3.02	4.55	0.61 %	**
smsa3120	2.06	0.44	2.06	4.65	0.56 %	**
smsa3160	1.68	0.36	1.68	4.60	0.58 %	**
smsa3280	2.85	1.31	2.85	2.18	8.09 %	.
smsa3320	0.31	0.19	0.31	1.61	16.93 %	
smsa3360	0.86	0.14	0.86	6.07	0.18 %	**
smsa3480	2.59	0.28	2.59	9.22	0.03 %	***
smsa3560	0.79	0.46	0.79	1.74	14.23 %	
smsa3600	1.47	0.25	1.47	5.83	0.21 %	**
smsa3640	2.18	0.21	2.18	10.54	0.01 %	***
smsa3660	2.78	1.30	2.78	2.14	8.50 %	.
smsa3760	1.99	0.21	1.99	9.42	0.02 %	***

Coefficients	Estimate	Std. Error	Marginal Effect	t value	Pr(>|t|)	Flag†
smsa3840	3.37	0.76	3.37	4.40	0.70 %	**
smsa3965	3.20	0.86	3.20	3.73	1.35 %	*
smsa4040	2.81	1.29	2.81	2.17	8.22 %	.
smsa4120	1.37	0.25	1.37	5.45	0.28 %	**
smsa4160	3.13	1.29	3.13	2.43	5.95 %	.
smsa4280	3.60	1.28	3.60	2.81	3.76 %	*
smsa4400	1.57	0.46	1.57	3.42	1.89 %	*
smsa4480	1.46	0.09	1.46	16.02	0.00 %	***
smsa4720	2.36	0.70	2.36	3.37	1.99 %	*
smsa4880	-0.30	0.25	-0.30	-1.21	27.98 %	
smsa4900	2.02	0.66	2.02	3.05	2.84 %	*
smsa4920	1.53	0.26	1.53	5.78	0.22 %	**
smsa5000	0.32	0.12	0.32	2.60	4.85 %	*
smsa5015	2.08	0.32	2.08	6.60	0.12 %	**
smsa5080	2.81	0.32	2.81	8.87	0.03 %	***
smsa5120	2.06	0.20	2.06	10.35	0.01 %	***
smsa5160	1.32	0.45	1.32	2.93	3.27 %	*
smsa5170	2.51	0.69	2.51	3.62	1.52 %	*
smsa5190	2.52	0.49	2.52	5.09	0.38 %	**
smsa5240	0.94	0.39	0.94	2.38	6.34 %	.
smsa5360	1.82	0.33	1.82	5.58	0.26 %	**
smsa5380	1.91	0.19	1.91	10.17	0.02 %	***
smsa5480	2.79	0.67	2.79	4.18	0.86 %	**
smsa5560	0.93	0.18	0.93	5.22	0.34 %	**
smsa5600	1.72	0.10	1.72	17.30	0.00 %	***
smsa5640	1.94	0.21	1.94	9.31	0.02 %	***
smsa5720	2.45	0.26	2.45	9.41	0.02 %	***
smsa5775	1.95	0.19	1.95	10.47	0.01 %	***
smsa5880	2.40	0.24	2.40	9.81	0.02 %	***
smsa5920	1.94	0.34	1.94	5.73	0.23 %	**
smsa5960	1.74	0.33	1.74	5.31	0.32 %	**
smsa6000	1.66	0.32	1.66	5.24	0.33 %	**
smsa6080	1.58	0.64	1.58	2.47	5.66 %	.
smsa6120	2.41	0.66	2.41	3.62	1.52 %	*
smsa6160	2.41	0.14	2.41	17.42	0.00 %	***
smsa6200	0.80	0.13	0.80	6.10	0.17 %	**
smsa6280	1.91	0.19	1.91	9.85	0.02 %	***
smsa6440	2.25	0.15	2.25	14.78	0.00 %	***
smsa6480	2.01	0.27	2.01	7.44	0.07 %	***
smsa6640	1.33	0.30	1.33	4.41	0.70 %	**
smsa6780	1.39	0.16	1.39	8.85	0.03 %	***

Coefficients	Estimate	Std. Error	Marginal Effect	t value	Pr(>|t|)	Flag†
smsa6840	3.03	0.52	3.03	5.86	0.21 %	**
smsa6920	1.83	0.22	1.83	8.26	0.04 %	***
smsa7040	2.57	0.22	2.57	11.80	0.01 %	***
smsa7160	1.41	0.24	1.41	5.92	0.20 %	**
smsa7240	0.74	0.17	0.74	4.29	0.78 %	**
smsa7320	1.41	0.14	1.41	9.74	0.02 %	***
smsa7360	1.52	0.18	1.52	8.56	0.04 %	***
smsa7400	1.76	0.23	1.76	7.59	0.06 %	***
smsa7480	3.30	1.30	3.30	2.55	5.14 %	.
smsa7500	1.69	0.75	1.69	2.26	7.36 %	.
smsa7510	0.23	0.43	0.23	0.53	61.64 %	
smsa7560	1.52	0.29	1.52	5.34	0.31 %	**
smsa7600	1.99	0.23	1.99	8.69	0.03 %	***
smsa7680	0.49	0.34	0.49	1.44	21.05 %	
smsa7840	0.89	0.38	0.89	2.36	6.51 %	.
smsa8000	4.24	1.32	4.24	3.21	2.37 %	*
smsa8120	1.58	0.42	1.58	3.72	1.37 %	*
smsa8160	1.89	0.50	1.89	3.79	1.28 %	*
smsa8200	1.59	0.41	1.59	3.82	1.23 %	*
smsa8280	1.48	0.20	1.48	7.57	0.06 %	***
smsa8400	1.72	0.31	1.72	5.46	0.28 %	**
smsa8480	5.94	1.33	5.94	4.46	0.67 %	**
smsa8520	1.41	0.27	1.41	5.26	0.33 %	**
smsa8560	2.34	0.49	2.34	4.81	0.48 %	**
smsa8720	1.17	0.58	1.17	2.01	10.04 %	
smsa8840	2.49	0.24	2.49	10.56	0.01 %	***
smsa8960	0.73	0.19	0.73	3.82	1.24 %	*
smsa9040	1.75	0.37	1.75	4.75	0.51 %	**
smsa9240	2.40	0.59	2.40	4.09	0.94 %	**
smsa9320	2.00	0.38	2.00	5.24	0.34 %	**
smsar.1	2.02	0.11	2.02	18.57	0.00 %	***
smsar.2	1.76	0.09	1.76	19.80	0.00 %	***
smsar.3	0.99	0.09	0.99	11.08	0.01 %	***
smsar.4	1.33	0.10	1.33	13.53	0.00 %	***

† Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1