Environmental Justice:
Evidence from Superfund Cleanup Durations[*]

Martin Burda[†] Matthew Harding[‡]

September 4, 2013

Abstract

This paper investigates the extent to which cleanup durations at Superfund sites reflect demographic biases incongruent with the principles of Environmental Justice. We argue that the duration of cleanup, conditional on a large number of site characteristics, should be independent of the race and income profile of the neighborhood in which the site is located. Since the demographic composition of a neighborhood changes during the cleanup process, we explore whether cleanup durations are related to neighborhood demographics recorded at the time when the cleanup is initiated. We estimate a semiparametric Bayesian proportional hazard model, which also allows for unobserved site specific heterogeneity, and find that sites located in black, urban and lower educated neighborhoods were discriminated against at the beginning of the program but that the degree of bias diminished over time. Executive Order 12898 of 1994 appears to have re-prioritized resources for the faster cleanup of sites located in less wealthy neighborhoods. We do not find that the litigation process is an impediment in the cleanup process, and support the notion that community involvement plays an important role.

JEL: Q53, Q58, C41

Keywords: Environmental Justice, Superfund, semiparametric Bayesian duration analysis

---

[*] We would like to thank the participants of the Conference on Bayesian Methods in Microeconometric Modeling at California Institute of Technology 2013 and the workshop on the Identification of Causal Effects in Environmental and Energy Economics at the University of Tennessee, Knoxville, 2013 for their insightful comments and suggestions. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and it was supported by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC: www.sshrc-crsh.gc.ca).

[†] Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada; Phone: (416) 978-4479; Email: `martin.burda@utoronto.ca`

[‡] Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: `mch@stanford.edu`

1. Introduction

This paper investigates the extent to which the cleanup process of toxic waste sites, known as Superfund sites, over the last 30 years was implemented in a fair way without inherent demographic biases. The Environmental Protection Agency (EPA) defines *Environmental Justice* as “the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies”. Environmental Justice considerations were formally established in 1994 when President Bill Clinton signed Executive Order 12898 which aimed to prevent discrimination in the implementation of environmental protection policies.

Evaluating Environmental Justice presents substantial challenges due to the inherent selection of the location of productive activity and residential sorting decisions taken over a long period of time. These may lead to the spurious correlation between neighborhood demographics and the presence of a hazardous waste site. This papers takes a novel identification approach to evaluating Environmental Justice claims. We analyze separate milestones in the cleanup process conditional on a large set of site characteristics (both observable and unobservable) and investigate whether the resulting *duration* of cleanup was in any way influenced by the demographic characteristics of the affected population. Since cleanups take many years to complete, we expect neighborhood demographics to also change as a result of the cleanup process itself. We avoid this potential source of endogeneity by relating the duration of the cleanups to neighborhood demographics at the very beginning of the cleanup process. This allows us to treat the factors driving the cleanup process as pre-determined with respect to the cleanup duration.

Our identification strategy requires us to model the cleanup duration conditional on a large set of observed and unobserved site characteristics. In spite of the richness of our data, which describes the nature of the contamination at a given site in detail, it is not possible to account for all site specific features which may influence the duration of the cleanup process. We therefore rely on a state of the art econometric model that accounts for the presence of unobserved nonparametrically distributed site specific effects. This added flexibility helps diminish potential biases due to model misspecification.

We further evaluate the extent to which demographic biases may have changed over time and in particular the degree to which the 1994 legislative change, which emphasized Environmental Justice considerations, altered the way Superfund cleanups are conducted. We find that sites located in black, urban, lower educated communities were discriminated against at the beginning of the Superfund program in the early 1980s. The degree of bias does diminish over time though and the emphasis placed on Environmental Justice after 1994 lead to faster cleanup times for Superfund sites located in poor neighborhoods. After the cleanup is completed, the time to return a site to general use depends almost exclusively on the economic health of the neighborhood.

We also investigate whether the observed demographic or economic biases may in fact reflect different aspects of the bargaining process between the government, the responsible parties and the local community. We do not find evidence that the Superfund litigation process is delaying Superfund cleanups. We do however find that community involvement plays an important role in the cleanup duration.

Various aspects of Superfund sites have been under scrutiny in the previous academic literature. Environmental Justice concerns were initially introduced by a number of correlation based studies which documented the presence of a relationship between the location of hazardous waste sites and the demographic composition of the adjacent neighborhoods (United Church of Christ (UCC) 1987). While considerable disagreement exists regarding how best to define a neighborhood, a number studies have documented the presence of racial and income inequalities in the geographic location of Superfund sites (Stretsky and Hogan 1998, Smith 2009, Sigman and Stafford 2010). A related strand of the literature investigates the process through which hazardous waste sites are designated as Superfund sites and finds that sites located in communities with a higher percentage of minorities are less likely to be listed on the National Priorities List (NPL) thereby delaying the cleanup process (Anderton, Oakes, and Egan 1997). It is not clear however to what extent the resulting biases documented in both strands of the literature reflect actual biases or the influence of unobserved factors that initially determined the nonrandom distribution of production activity and hazardous waste location in the country. Wolverton (2009) shows that when plant locations are associated with current demographic characteristics, both race and income predict plant locations. However, when plant locations are associated with demographic characteristics at the time of the siting race is no longer a significant predictor.

Limited attention has been given to the duration of cleanup at Superfund sites. Beider (1994) uses a survey of EPA site managers to investigate the main reasons for the long cleanup durations and concludes that the primary reasons are the inherent difficulty of cleanup (i.e. the extent and nature of the contamination process) and the associated legal process which may involve many parties. Sigman (2001) is the only study we are aware of which employs a formal econometric model for Superfund cleanup durations. It employs a Weibull model which relies on a parametric specification with limited ability to account for unobserved heterogeneity. The paper finds that the extent of contamination and the nature of the liable parties explain the durations. However, higher income communities were found to have longer cleanup durations.

The benefits of cleanup are substantial. Currie, Greenstone, and Moretti (2011) report that Superfund cleanups reduce the incidence of congenital anomalies in newly born babies by up to 25%. In general though, it is difficult to quantify the cleanup influence on human health precisely and incorporate it in a traditional cost benefit analysis (Hamilton and Viscusi 1999). For example, measuring human health benefits in terms of the number of cancer cases avoided requires assumptions on any number of behavioral and environmental confounders over a life time. One of the difficulties also comes from the fact that we often have to rely on indirect approaches, e.g. by looking at the impact of Superfunds on the housing market, which may conflate the true benefit of Superfund cleanups with informational or reputational considerations (Gayer, Hamilton, and Viscusi 2000, Greenstone and Gallagher 2008).

This paper proceeds as follows. In Section 2 we discuss the cleanup process and distinguish between the various milestones in the cleanup of a Superfund site. Section 3 introduces the available data. We elaborate on our approach to identifying the presence of demographic biases that may be incongruent with Environmental Justice considerations. Since our identification strategy requires the estimation of a complex econometric model, we also discuss our estimation strategy in detail. Section 4 presents the main empirical results, while Section 5 explores the robustness of these results to alternative explanations based on the degree of bargaining power between the different parties involved in the cleanup process. In particular we investigate the role of litigation and community involvement activities. Section 6 concludes.

2. **The Superfund Cleanup Process**

Over the years policy makers have become increasingly aware of both the need to regulate dangerous substances and also to address the existing stock of hazardous waste sites. The most well-known effort to clean up hazardous waste sites, commonly known as Superfund, provides broad federal authority to the Environmental Protection Agency (EPA) to clean up or compel the responsible parties to clean up the most hazardous of these sites.

Waste is an inevitable part of the production process. The 2010 census counted more than 5.7 million firms and over 7.3 million establishments. It has been estimated that over 600,000 establishments are currently generating waste which can be classified as hazardous to human health (Sigman and Stafford 2010). This includes many types of substances which are known to be toxic, ignitable, radioactive, or in some other fashion present a real danger to the nearby population. In addition there are many hazardous waste sites resulting from production activity or inappropriate storage in past decades which resulted in soil and water contamination, such as abandoned factories and warehouses, landfills and military installations.

In this paper we explicitly focus on and model the durations of two main stages in the Superfund cleanup process, which we will briefly review.[4] To become a Superfund site, a hazardous waste site must go through an evaluation process. This process consists of discovery, evaluation, and nomination of contamination sites to the Superfund National Priorities List (NPL) as defined in the Comprehensive Environmental Remediation, Compensation, and Liabilities Act (CERCLA).

The Superfund process begins with the discovery of a Superfund site or notification to EPA of the possible release of hazardous substances. Site discovery can be initiated by a number of different parties, including citizens, businesses, State or local government and EPA regional offices. Once a site has been discovered, it is entered into the Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS). The site is then evaluated to determine whether it meets the qualifications for listing on the NPL.

The first step in this evaluation is a Preliminary Assessment (PA) to determine if the site has the potential to qualify for the NPL. This is a limited screening investigation to distinguish sites that pose little to no potential threat. During this stage, readily available information about the site is collected. If it is determined that the site indeed poses little to no threat, then the process stops here. If instead the evaluation determines that the site may pose a threat to human health or the environment and therefore may qualify for the NPL, Site Inspection (SI) will commence. At this point environmental and waste related data is collected and analyzed. This data is then used to determine if the site qualifies for the NPL. The data will also be used to score the site based on the Hazard Ranking System (HRS). The HRS is a quantitative based tool to assess the relative degree of risk to the environment and human health by a potential or actual release of hazardous substances.

The proposal to list the site on the NPL and the HRS package is placed on the Federal Register. After a preliminary investigation, if the site is still found to qualify for NPL, then it will be placed on the NPL and the remedial process will begin. For our purpose we consider the NPL listing date as the initial starting point of the cleanup process.

Once a site is listed on the NPL the first stage of the cleanup process, the “remedial program” begins. First, a detailed examination of the site ensues which determines the precise nature of the contamination and the technical requirements for cleaning up the site. At this stage the EPA is required to solicit public opinion in the evaluation of the various cleanup options. Once this evaluation is completed a Record of Decision (ROD) is issued which describes the precise nature of the cleanup process to be implemented and the nature of the eventual cleanup target. After this, the various actions listed in the ROD commence and it will normally take years for the actions to

---

[4] More detailed information can be found on the EPA website: `http://www.epa.gov/superfund/`

be implemented. This is not unexpected given the technical challenges encountered in the process of removing the hazardous substances involved and containing or cleaning the contamination of surrounding soil and water. The *first milestone* in the cleanup process consists of the date when a site is labeled as "construction complete". This indicates that all physical or engineering tasks have been completed and both immediate and long term threats have been addressed. Note that construction complete does not mean that all threats have been neutralized and the cleanup goals have been achieved. For example, it is possible for the source of the contamination to have been completely removed but the surrounding media to remain toxic and thus not ready for being returned to general use.

The post construction complete phase may involve a number of different activities necessary for achieving the ultimate clean up goals. For example, ground water restoration may require prolonged ongoing treatment. Other hazardous sites may require ongoing monitoring and restricted access for many years after the engineering effort has ceased. This process is subject to regular reviews until it is determined that all cleanup goals have been met and no further action is required. At that point, the site reaches the *second milestone* in the cleanup process, when it is "deleted" from the NPL. Depending on the nature of the site it may then be reused or redeveloped for a new purpose.

In this paper we use two different measures of the cleanup durations in the analysis. Since the processes involved for reaching the two different milestones are different we expect each measure to be informative in its own right. Therefore, we do not restrict the model parameters across the duration types and estimate separate models for each duration. The durations are as follows: (1) the duration between a site being listed on NPL and the construction being completed at the site; (2) the duration between the construction being completed and the site being deleted from NPL.

3. **Data and Empirical Strategy**

In this paper we use data obtained from the EPA on all sites listed on NPL between 1983 and the end of 2010. In Figure 1 we plot the histograms for the two durations. Many of the observations are censored and this feature will need to be accounted for in the estimation. The mean values for the two durations are 13.8 years, and 9.0 years respectively. The first milestone is reached by most sites within 20 years. Sites for which the construction complete process has not been reached within 20 years are substantially more likely to be censored by the end of 2010. In contrast if the second milestone is reached, then it is reached for most sites within 5 years, indicating that the cleanup goals are achieved relatively soon after construction is completed. Nevertheless, for a substantial number of the sites this milestone is not reached indicating that only a fraction of the sites have been returned to general use so far.

For each site we observe its location and also a very comprehensive description of the form of contamination at that site. In particular we see the nature of the contaminated media (debris, groundwater, sediment, surface water, or waste) and the type of contaminants from acids to radioactive substances and volatile organic compounds (VOC). We believe this to be both an accurate and comprehensive description of the challenges encountered at the site and the degree of difficulty to clean it up. In particular note that many sites have both varied contaminated media and numerous contaminants that need to be addressed. The presence of a type of contaminated media or contaminant at a site is recorded in the form of an indicator variable. In Table 1 we report the means and standard deviations of the contaminants and contaminated media of all sites listed for each of the decades 1980s, 1990s, and 2000s. We notice a substantial degree of heterogeneity both across contaminants and decades. In particular the presence and extent of contamination appears to be decreasing over time. This is consistent with the notion that the most hazardous and challenging sites were detected in the early years of the Superfund program and that advances in regulation have reduced, although not eliminated, the occurrence of new hazardous waste. No new sites were listed during the 2000s that were contaminated with radioactive materials or where the contaminated media consisted of debris (often in the form of building remains contaminated with asbestos).

During the preliminary assessment and site inspection, each site is allocated an HRS score on the EPA Hazard Ranking System. The score is computed by aggregating along a number of different dimensions such as the characteristics of the waste (toxicity and quantity), the extent of hazardous waste released or expected to be released into the environment, the intensity with which people may be affected, and the degree to which ground water, surface water, soil and air have been exposed. The HRS score is designed to capture the nature of the site's hazard used to decide whether the site should be placed on the NPL. Note however, that according to the EPA[5] the HRS is *not* sufficient to prioritize the cleanup process at a Superfund site. In particular, given resource constraints, a high HRS score does not imply the reallocation of funds from existing cleanups already in process. Thus, while HRS is correlated with the degree to which a site is hazardous and it plays an important role in the placement of a site on the NPL, we expect it to be only weakly related to the cleanup duration itself. Table 1 reveals a small increase in the HRS scores at listing for sites over the three decades.[6]

A crucial component for determining the cleanup strategy consists in compiling the Record of Decision (ROD). The ROD presents details on the planned cleanup implementation. The costs

---

[5] `http://www.epa.gov/superfund/programs/npl_hrs/hrsint.htm`

[6] Note that this does not mean that sites listed later are more contaminated. A less contaminated site can have a larger score if the contamination presents a risk to a larger population.

recorded in the RODs are projected for the alternative selected from many possible options. These include capital costs, transaction costs, and operation and management costs. We have individually reviewed the RODs for all NPL sites and extracted from them a measure of the estimated present value of the cleanup costs at an assumed interest rate of 7%. Table 1 does not indicate any consistent trend in the costs associated with the cleanup process over time.

Note however that about 5% of the sites on NPL have a recorded costs of 0. In this case the selected alternative was "no further action". This could happen due to two possible reasons: (1) upon further consideration it was determined that there was no threat to human life or the environment, and (2) an immediate threat required removal action and by the time the rest of the procedures (everything up to the ROD) were completed, no further action was needed. We consider these sites to be different from other sites and assign them a separate indicator variable.

For each site we use the site location to obtain the population demographics in the zip code in which the Superfund site is located at the time of listing. We use the 1980 census to capture the demographics for a site listed between 1981 and 1989, and similarly for other decades. We record the median household income, and the fractions of the population which are college educated, black, and urban. Furthermore we record the fractions of the population by age. Table 1 shows that the demographic composition of the neighborhoods in which the hazardous sites were located varied with the time when the site was listed on the NPL. Sites listed earlier were more likely to be located in affluent, white neighborhoods, while sites listed later were more likely to be located in urban neighborhoods with a higher percentage of college educated residents.[7] Sites listed earlier were also more likely to be located in neighborhoods with younger residents. We condition on the event of the sites' NPL listing and control for their demographic characteristics.

The Superfund discovery is a distinct process beyond the scope of our analysis. Recall that earlier studies have found that sites located in neighborhoods with a higher percentage of minority residents seem less likely to be placed on the NPL (Anderton, Oakes, and Egan 1997). It would be difficult to convincingly model the discovery process itself, as the location and nature of the contamination were determined in most cases decades before the discovery process was initiated.

3.1. **Identification**

This paper focuses on evaluating the extent to which biases based on demographic characteristics such as income and race may be affecting the cleanup durations at Superfund sites in violation of the principles of Environmental Justice. The main assumption, which drives the identification

---

[7] Note that in the tables for this paper we use the name Bachelor+ to refer to the percentage of the population which has obtained at least a BA degree.

of our model, is that the duration of cleanup is based purely on a rational cost-benefit analysis which depends on a wide range of site specific factors (both observed and unobserved by the econometrician) and a common baseline hazard which reflects macroeconomic trends and potentially the variation in the Superfund budget. Departures from the cost-benefit framework, e.g. in the form of faster cleanups observed in wealthier neighborhoods, indicate the presence of demographic biases. Our approach to identification is similar to that chosen by Viscusi and Hamilton (1999), who interpret departures from the cost-benefit analysis in the decisions taken by regulators regarding the chemical cleanup targets at Superfund sites as evidence of departures from rationality, behavioral biases, and risk misperceptions.

Our framework assumes that the following set of factors provides a comprehensive model explaining the durations between the different cleanup stages:

(1) the set of contaminants recorded at each site;
(2) the set of contaminated media at each site;
(3) the HRS score for each site;
(4) the engineering estimate of the cost of cleanup based on the original ROD;
(5) the time when a site was listed;
(6) information on the parties involved in litigation;
(7) information on the degree to which the community was involved in the cleanup decision process;
(8) an aggregate trend capturing the impact of the macroeconomy or the Superfund budget common to all sites;
(9) a site specific time invariant random effect.

In order to test for the presence of biases we augment this model with demographic variables *pre-determined* for each site at the time of listing. This avoids the potential endogenous feedback between the duration of the cleanup and subsequent demographic and environmental changes. In Table 2 we explore the extent to which neighborhood demographics change after a site is designated a Superfund site. Income, in particular, declines sharply during the cleanup period. This effect is not limited to the early years after a cleanup begins, which may be driven by residents leaving an area once they become aware of the presence of a Superfund site in their neighborhood. Median income continues to decline even after 20 years of cleanup activities. If we were to correlate the duration of cleanup with the change in the demographic composition of the neighborhood we would find a large negative correlation between the duration of cleanup and the change in median income. It would however be misleading to interpret this correlation as implying that wealthy neighborhoods are cleaned up faster, since it is likely that the composition of the a neighborhood changes as wealthier households leave a neighborhood with a Superfund site that is being cleaned up. Therefore, we only use the demographic composition of a site at the time of its listing to explain the subsequent cleanup duration. In order to test for possible violations of Environmental Justice we then test for the significance of the demographic variables at the time when a site was listed on NPL. This approach is similar to that used by Wolverton (2009) who investigates the relationship between firm locations and neighborhood demographics by focusing on the demographic composition of the neighborhood at the time when the location decision was made.

Our econometric model allows for the the cleanup duration to also depend on a site specific effect, which our Bayesian hierarchical model allows to be correlated with the observed site attributes. Our estimation procedure will estimate the distribution of these effects in the sample. The rationale behind including a site specific effect is that in spite of the richness of our data, which captures many of the observed site characteristics, it is nevertheless possible that not all features of the site which are relevant for the cleanup process have been recorded and which may lead to an omitted variables bias. Consider for example the period of time a site was contaminated before the cleanup process was initiated. We do not observe this in the data and it may be correlated with the severity of the contamination. Furthermore, sites that have been contaminated for a longer period of time may be inherently more difficult to clean up or may require more intensive and time consuming engineering processes. This variable may also be correlated with neighborhood characteristics, since the timing of the location could have been driven by the latter. Below we introduce the econometric model and its technical assumptions.

3.2. **Econometric Model**

In order to quantify the degree to which the duration of the cleanup process is biased by the demographic characteristics of the neighborhoods in which the Superfund sites are located, we develop a state of the art econometric model of the duration between the different milestones in the Superfund cleanup process. The model builds on the recent work of Burda, Harding, and Hausman (2012) (BHH) who introduce a flexible semiparametric Bayesian proportional hazard duration model. The model allows for the presence of time variant or invariant observables but also models the baseline hazard and the site specific unobserved heterogeneity nonparametrically. While BHH devised their model for interval outcome data whereby only a general time period of the duration outcome was observed, here we alter their model to make use of the exact timing of the duration with point-in-time outcomes.

Denote by t_i the point in time elapsed when a site i was observed to exit from a given state into another state. Define the hazard rate λ_{it} as the failure rate at time t conditional upon survival to time t, $\lambda_{it} = \lim_{\delta \to 0} Pr(t < t_i < t + \delta)/\delta$ and denote the integrated hazard by:

$$\Lambda_{it} = \int_0^t \lambda_{i\tau} d\tau \tag{3.1}$$

The survivor function S_{it} and the distribution function F_{it} of t are defined as

$$S_{it} = \exp(-\Lambda_{it}) \tag{3.2}$$

$$F_{it} = 1 - S_{it} \tag{3.3}$$

Hence, the conditional density function of exit at t is given by

$$\begin{aligned} f_{it} &= F'_{it} \\ &= -S'_{it} \\ &= \exp(-\Lambda_{it}) \lambda_{it} \\ &= S_{it}\lambda_{it} \end{aligned} \tag{3.4}$$

which forms the contribution to the conditional likelihood function for non-censored data. For observations censored at time T, all we know under non-informative censoring is that the lifetime exceeds T. The probability of this event, and therefore its contribution to the likelihood is

$$\begin{aligned} P(t_i > T) &= 1 - F_{iT} \\ &= S_{iT} \end{aligned} \tag{3.5}$$

The likelihood terms (3.4) and (3.5) can be written as the single expression

$$L_i(t_i) = S_{it}\lambda_{it}^{d_i} \tag{3.6}$$

where d_i is a censoring indicator variable taking the value of 1 if $t_i \leq T$, or the value of 0 if $t_i > T$, in which case t_i is set to equal T in (3.6).

ASSUMPTION (1). *The data* $\{t_i\}_{i=1}^N$ *consists of single spells censored at time* T *and drawn from a single risk process.*

ASSUMPTION (2). *The hazard rate is parameterized as*

$$\lambda_{it} = \lambda_{0t} \exp(X_{it}\beta + V_i) \tag{3.7}$$

where λ_{0t} *is the baseline hazard,* X_{it} *are observed covariates that are allowed to vary over time,* β *are model parameters, and* V_i *is an unobserved heterogeneity component.*

ASSUMPTION (3). *The baseline hazard* λ_{0t} *and the values of the covariates* X_{it} *are constant over time intervals* $[t_{j-1}, t_j)$ *for* $j = 1, \ldots, J$.

Assumptions 1 and 2 are common in the literature. Assumption 3 is based on Han and Hausman (1990). Given Assumption 3, we can consider the integrated baseline hazard in the form

$$\mu_{0j} = \int_{t_{j-1}}^{t_j} \lambda_{0\tau} d\tau \tag{3.8}$$

where we denote the vector $(\mu_{01,} \ldots, \mu_{0J})$ by μ_0.

Denote by $[\underline{t_i}, \overline{t_i}) \ni t_i$ the time interval during which i's exit occurred, with endpoints $\underline{t_i} \in \{t_0, \ldots, t_{J-1}\}$ and $\overline{t_i} \in \{t_1, \ldots, t_J\}$ with $\{t_j\}_{j=0}^{J}$ as defined in Assumption 3. Define the variable

$$\iota_{ij} = \begin{cases} 1 \text{ if } t_i \notin [t_{j-1}, t_j) \\ \left(t_i - \underline{t_i}\right) / (\overline{t_i} - \underline{t_i}) \text{ if } t_i \in [t_{j-1}, t_j) = [\underline{t_i}, \overline{t_i}) \end{cases} \tag{3.9}$$

Using Assumptions 1–3 and the notation in (3.9), the conditional likelihood (3.6) can now be rewritten as

$$L_i(t_i; V_i) = \exp\left(- \sum_{j=1}^{\overline{t_i}} \iota_{ij} \mu_{0j} \exp\left(X_{ij}\beta + V_i\right) \right) \left\{ \left(\mu_{0\overline{t_i}} / (\overline{t_i} - \underline{t_i})\right) \exp(X_{i\overline{t_i}}\beta + V_i) \right\}^{d_i} \tag{3.10}$$

3.3. **Parametric Heterogeneity**

ASSUMPTION (4)**.** *Let*

$$v_i \equiv \exp(V_i) \sim \mathcal{G}(v)$$

where $\mathcal{G}(v)$ *is a probability distribution function with density* $g(v)$*.*

Using Assumption 4, denote by tilde the part of the hazard without the heterogeneity term:

$$\lambda_{ij} = v_i \widetilde{\lambda}_{ij} \tag{3.11}$$

where, from Assumption 3 and (3.7),

$$\widetilde{\lambda}_{ij} = (\mu_{0j} / (t_j - t_{j-1})) \exp(X_{ij}\beta)$$

Hence, at the time of exit t_i,

$$\widetilde{\lambda}_{it_i} = \left(\mu_{0\overline{t_i}} / (\overline{t_i} - \underline{t_i})\right) \exp(X_{i\overline{t_i}}\beta) \tag{3.12}$$

Similarly, using Assumption 4, let

$$\Lambda_{it} = v_i \widetilde{\Lambda}_{it} \tag{3.13}$$

where, from (3.1),

$$\widetilde{\Lambda}_{it} = \int_0^t \lambda_{0\tau} \exp\left(X_{i\tau}\beta\right) d\tau$$

Due to Assumption 3, (3.8), (3.9), and (3.13), at the time of exit t_i,

$$\widetilde{\Lambda}_{it_i} = \sum_{j=1}^{\overline{t_i}} \iota_{ij} \left(\mu_{0j}/(t_j - t_{j-1})\right) \exp\left(X_{ij}\beta\right) \tag{3.14}$$

If v is a random variable with probability density function $g(v)$ then the Laplace transform of $g(v)$ evaluated at $s \in \mathbb{R}$ is defined as

$$\begin{aligned} \mathcal{L}(s) &\equiv \int \exp(-vs) g(v) dv \\ &= E_v[\exp(-vs)] \end{aligned} \tag{3.15}$$

and its r-th derivative is

$$\mathcal{L}^{(r)}(s) \equiv (-1)^r \int v^r \exp(-vs) g(v) dv \tag{3.16}$$

Using (3.11), (3.13), and (3.15), the expectation of the survival function can be linked to the Laplace transform of the integrated hazard function (Hougaard, 2000) as

$$E_v\left[S_{it}\right] = \mathcal{L}(\widetilde{\Lambda}_{it}) \tag{3.17}$$

which forms the expected likelihood for censored observations.

For uncensored observations, collecting (3.11), (3.12), (3.13), and (3.14) in (3.10), yields

$$L_i(t_i; V_i, d_i = 1) = \exp\left(-v_i \widetilde{\Lambda}_{it_i}\right) v_i \widetilde{\lambda}_{it_i}$$

Taking expectations and using (3.16) we obtain

$$\begin{aligned} E_{v_i}\left[L_i(t_i; V_i)\right] &= E_{v_i}\left[\exp\left(-v_i \widetilde{\Lambda}_{it_i}\right) v_i \widetilde{\lambda}_{it_i}\right] \\ &= \widetilde{\lambda}_{it_i} E_{v_i}\left[\exp\left(-v_i \widetilde{\Lambda}_{it_i}\right) v_i\right] \\ &= -\widetilde{\lambda}_{it_i} \mathcal{L}^{(1)}(\widetilde{\Lambda}_{it_i}) \end{aligned} \tag{3.18}$$

The expected likelihood terms (3.17) and (3.18) are summarized in the following Result:

RESULT 1. *The expectation of the likelihood (3.6) with respect to unobserved heterogeneity, distributed according to a generic probability measure as given by Assumption 4, is for uncensored observations*

$$E_{v_i}\left[L_i(t_i; V_i, d_i = 1)\right] = -\widetilde{\lambda}_{it_i} \mathcal{L}^{(1)}(\widetilde{\Lambda}_{it_i}) \tag{3.19}$$

and for censored observations

$$E_{v_i}\left[L_i(t_i; V_i, d_i = 0)\right] = \mathcal{L}(\widetilde{\Lambda}_{iT}) \tag{3.20}$$

Since the site heterogeneity term v_i defined in Assumption 4 is non-negative, a suitable family of distributions $\mathcal{G}(v)$ with support over $[0,\infty)$ and tractable closed-form Laplace transforms is Generalized Inverse Gaussian (GIG) class of distributions, whose special case, among others, is the gamma distribution popular in duration analysis.

ASSUMPTION (5). *The unobserved heterogeneity term v_i is distributed according to the Generalized Inverse Gaussian distribution,*

$$\mathcal{G}(v) = \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta)$$

The GIG has the density

$$g^{GIG}(v;\kappa,\varphi,\theta) = \frac{2^{\kappa-1}}{K_\kappa(\varphi)}\frac{\theta}{\varphi^\kappa}(\theta v)^{\kappa-1}\exp\left\{-\theta v - \frac{\varphi^2}{4\theta v}\right\} \tag{3.21}$$

for $\varphi,\theta > 0$, $\kappa \in \mathbb{R}$, where $K_\kappa(\varphi)$ is the modified Bessel function of the second kind of order κ evaluated at φ (Hougaard, 2000). The GIG Laplace transform is given by

$$\mathcal{L}^{GIG}(s;\kappa,\varphi,\theta) = (1+s/\theta)^{-\kappa/2}\,\frac{K_\kappa\left(\varphi\,(1+s/\theta)^{1/2}\right)}{K_\kappa(\varphi)} \tag{3.22}$$

and its derivatives by

$$\mathcal{L}^{(r)GIG}(s) = (-1)^r\frac{K_{\kappa+r}\left(\varphi\,(1+s/\theta)^{1/2}\right)}{K_\kappa(\varphi)}\left(\frac{\varphi}{2\theta}\right)^r(1+s/\theta)^{-(\kappa+r)/2} \tag{3.23}$$

The GIG family includes as special cases the gamma distribution for $\varphi = 0$, the Inverse gamma distribution for $\theta = 0$, and the Inverse Gaussian distribution for $\kappa = -\frac{1}{2}$, among others.

Application of the Laplace transform of the GIG distribution (3.22) and its derivatives (3.23) in Result 1 yields the following result:

RESULT 2. *Under the Assumptions 1–5,*

$$E_{v_i}\left[L_i(t_i;V_i,d_i=1)\right] = \frac{\varphi}{2}\frac{\widetilde{\lambda}_{it_i}}{\theta}\left(1+\frac{\widetilde{\Lambda}_{it_i}}{\theta}\right)^{-(\kappa+1)/2}\left[K_\kappa(\varphi)\right]^{-1}K_{\kappa+1}\left(\varphi\left(1+\frac{\widetilde{\Lambda}_{it_i}}{\theta}\right)^{1/2}\right) \tag{3.24}$$

and for the censored observations

$$E_{v_i}\left[L_i(t_i;V_i,d_i=0)\right] = \left(1+\frac{\widetilde{\Lambda}_{iT}}{\theta}\right)^{-\kappa/2}\left[K_\kappa(\varphi)\right]^{-1}K_\kappa\left(\varphi\left(1+\frac{\widetilde{\Lambda}_{iT}}{\theta}\right)^{1/2}\right) \tag{3.25}$$

A special case of the GIG distribution is the gamma distribution, obtained from the GIG density function (3.21) when $\varphi = 0$ and κ is restricted to the positive part of the real line.

The scale parameter θ has the feature that for any $c \in \mathbb{R}_+$, if $v \sim \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta)$ then $cv \sim \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta/c)$. Due to this property, c and hence its inverse $s \equiv c^{-1}$ are not separately identified from θ in the Laplace transform (3.22). Since all likelihood expressions are evaluated at $s = \widetilde{\Lambda}_{it}$ which is proportional to μ_{0j} for all j, as specified in (3.8), any change in θ only rescales the baseline hazard parameters μ_{0j}, leaving the likelihood unchanged. Hence, θ needs to be normalized to identify μ_{0j} by the moment restriction $E[v] = 1$.

3.4. **Flexible Heterogeneity**

We now depart from the parametric form of the unobserved heterogeneity and instead consider a nonparametric infinite mixture for the distribution of v_i, as formulated in the following assumption.

ASSUMPTION (6)**.** *The prior for v_i takes the form of the hierarchical model*

$$\begin{aligned} t_i &\sim F(v_i) \\ v_i|G &\sim G \\ G &\sim DP(G_0,\alpha) \\ \alpha &\sim \Gamma(a_0,b_0) \\ E[v_i] &= 1 \end{aligned}$$

In Assumption 6, G is a random probability measure distributed according to a Dirichlet Process (DP) prior (Hirano, 2002; Chib and Hamilton, 2002). The DP prior is indexed by two hyperparameters: a so-called baseline distribution G_0 that defines the "location" of the DP prior, and a positive scalar precision parameter α. The distribution G_0 may be viewed as the prior that would be used in a typical parametric analysis. The flexibility of the DP mixture model environment stems from allowing G to stochastically deviate from G_0. The precision parameter α determines the concentration of the prior for G around the DP prior location G_0 and thus measures the strength of belief in G_0. For large values of α, a sampled G is very likely to be close to G_0, and vice versa. Assumption 6 is then completed by specifying the baseline measure G_0 as follows:

ASSUMPTION (7)**.** *In Assumption 6,*

$$G_0 = \mathcal{G}^{GIG}(\kappa,\varphi,\theta) \tag{3.26}$$

Implementation of the GIG mixture model under Assumptions 1–3, 6, and 7 uses the probabilities (3.6), (3.24) and (3.25).

Under Assumptions 6 and 7, as a special limit case, putting all the prior probability on the baseline distribution G_0 by setting $\alpha \to \infty$ would result in forcing $G = G_0 = \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta)$ which yields a parametric model. Here we allow α and hence G to vary stochastically and the parametric benchmark specification is nested as a special case in our model.

3.5. **Marginal Effects**

One of the challenges of interpreting the economic significance of the empirical results lies in the inherent difficulty of computing marginal effects in this highly non-linear setting. Thus, while the estimated coefficients correctly capture the sign of the effect of interest it is non-trivial to translate the magnitude into an easily interpretable quantity. While we follow the established statistical practice of reporting the estimated coefficients, we also go a step further and use a simulation based approach to computing the economic significance of the statistically significant coefficients that are likely to be of particular interest to the reader.

There is no unique way of computing marginal effects in this type of non-linear model. We choose a simulation based approach which computes the average marginal effects for a discrete change in the variable of interest over the sample and using a large number of repeated draws from the distribution of unobserved heterogeneity. The economic significance of a coefficient is most easily interpretable in terms of time and we thus report the impact of a discrete change in the variable of interest as a fraction or multiple of 1 year of additional cleanup.

The expectation of a non-negative random variable t truncated at T is given by

$$E\left[t|t \leq T\right] = \frac{1}{F(T)} \int_0^T tf(t)dt \tag{3.27}$$

where $f(t)$ and $F(t)$ are pdf and cdf of t, respectively.

In our model where t denotes duration to cleanup,

$$\begin{aligned} f_i(t) &= \exp\left(-\Lambda_i(t)\right) \lambda_i(t) \\ F_i(t) &= 1 - \exp\left(-\Lambda_i(t)\right) \end{aligned}$$

Under the assumption of piece-wise constant baseline and covariates over time

$$E\left[t|t \leq T\right] = \frac{1}{1 - \exp\left(-\Lambda_{iT}\right)} \sum_{j=1}^{T} j \exp\left(-\Lambda_{ij}\right) \lambda_{ij} \tag{3.28}$$

where

$$\begin{aligned} \lambda_{ij} &= \mu_{0j} \exp\left(X_{ij}\beta + V_i\right) \\ \Lambda_{ij} &= \sum_{s=1}^{j} \lambda_{is} \end{aligned}$$

is the hazard and cumulative hazard, respectively.

For the effect of a δ change of X_{ijk} on $E\left[t|t \leq T\right]$, we evaluate

$$\Delta E\left[t|t \leq T\right] = E\left[t|t \leq T, X_{ijk} + \delta\right] - E\left[t|t \leq T, X_{ijk}\right]. \tag{3.29}$$

Since, the choice of δ is arbitrary for continuous variables, we follow Sigman (1998) and simulate the economic significance of a change of one standard deviation in the relevant covariate. The choice of truncation is also arbitrary and we truncate the simulated distribution at the latest date available in the sample. The simulations produce a number of outliers which need to be removed for meaningful results, which roughly correspond to values in the top and bottom 2% of the distribution. In the text below we report the economic significance of the main variables when describing the empirical results. More detailed tables are available from the authors.

4. Empirical Findings

Our current empirical framework allows us to investigate a number of important hypotheses regarding the main factors driving the cleanup durations between the two milestones in the cleanup process. It is important to note that throughout our identification strategy rules out the impact of sorting on the cleanup process. We then proceed to measure the extent to which the cleanup process was driven by cost-benefit factors associated with the engineering decisions regarding the technological aspects of the cleanup.

To the extent that demographic variables remain significant drivers of the cleanup durations we then proceed to investigate whether this reflects some form of direct discrimination or is perhaps a more indirect form of discrimination resulting from the differential bargaining ability of the different agents involved in the cleanup. In particular we differentiate between:

1. the role of the legal system and the bargaining power of the responsible parties,
2. the impact of general collective action as proxied by home ownership in the community,
3. a direct measure of Superfund related involvement as measured by EPA reported community activities.

4.1. Cleanup Durations

We consider a series of model specifications designed to estimate the factors determining the two cleanup durations of interest: the duration between listing and construction completion (LC), and the duration between construction completion and deletion (CD) from the NPL list. Recall that listing refers to the time when a site is listed on NPL, completion refers to the time when the remedial process has been completed, and deletion refers to the time when the site is removed from

NPL and returned to general use. These models capture our baseline identification approach and are developed to test a number of hypotheses of interest.

Our aim is to control for site characteristics (both observable and unobservable) and also for the demographic characteristics of the households potentially impacted by the site. Under our identification assumption we expect the presence of statistically significant coefficients on the demographic characteristics to be indicative of biases potentially incompatible with Environmental Justice considerations. In all specifications we model the conditional hazard rate for each site, yielding the probability that a site reaches the next milestone in the cleanup process. An estimated *negative* coefficient implies a lower probability of reaching the next milestone and a *slower* cleanup (longer cleanup duration).

In Table 3 we first estimate a simple duration model without unobserved heterogeneity which relates the two durations of interest, LC and CD, to neighborhood demographics only. These models are misspecified as a result of omitting a number of potentially important explanatory variables. Here, neighborhood demographics are strong predictors of the cleanup durations. Higher income, unemployment and the fraction of the population which is black are all associated with slower cleanup times. We then add observable site characteristics to the specification. These include engineering cost estimates and the description of the contaminants and contaminated media. If we now re-evaluate the relevance of the neighborhood demographics we find that their impact has been greatly diminished and most coefficients on the demographic variables become statistically insignificant.

In Table 4 we estimate the specification with both neighborhood demographics and observed site characteristics (columns 5-8 in Table 3) while also allowing for the presence of unobserved site specific effects. For each duration we estimate the corresponding model allowing for either a parametric specification or a nonparametric specification of the unobserved heterogeneity. While there are some noticeable differences, the models are comparable. From an econometric perspective, we consider the nonparametric model to be superior to the parametric one, in that the former nests the latter as a special case which may or may not be supported by the data evidence. This implies that in the nonparametric model the coefficient estimates of the demographic characteristics are likely to be more accurate and less confounded by the presence of unobserved site specific factors. Therefore, in all other tables, we will only report estimates derived from the nonparametric model.

First, consider the baseline model for the duration between listing and completion in Table 4. The impact of the HRS score is small, negative, and statistically significant. This is consistent with the EPA strategy of using the HRS scores to determine whether a site should be listed on NPL but not using the HRS scores directly to prioritize the cleanup activities, even though it reflects the extent to which a site is hazardous. The engineering cost estimates for the clean-up constitute a large and significant LC duration predictor. These costs are determined by the choice of remedy adopted and proxy for the complexity of the engineering process involved. We also include in our model an indicator for sites who have zero cost recorded in the documents available from the EPA. These are sites that were considered a priority and the cleanup was initiated immediately before an ROD was compiled because of the imminent danger to the population and the environment. The coefficient on this variable is an order of magnitude larger the one on the cost variable, reflecting the severity of the contamination at these sites.

The nature of contamination and the inherent technical difficulties involved in the cleanup process are major determinants of the cleanup durations. As we would expect sites containing metals, radioactive or PCB waste take longer to clean up. The contaminated media also represent an important factor. Sites where the waste takes the form of debris or waste which can be easily removed are much faster to clean up than sites where the sediment or soil is contaminated.

When considering the demographic variables we do not find statistical evidence that sites in minority neighborhood or low income neighborhoods are cleaned up slower. In fact we find that sites in wealthier neighborhoods are cleaned up slower but that sites located in neighborhoods with a large fraction of the population over 65 are cleaned up faster. In general we expect both wealthier and retired people to be more actively engaged in the construction decision process. Their incentives will vary however. Wealthy households are likely to prefer a comprehensive remedial process which will safeguard house prices by implementing more detailed and costly engineering approaches. On the other hand, older retired households may prefer a fast remedial process.

Let us now consider the corresponding models for the duration between completion and deletion. Sites with higher cleanup costs have longer durations. Contamination with metals, pesticides, and VOC impose additional challenges and extend the period it takes for the EPA to release a site for general use. Sites with contaminated groundwater are particularly challenging to clean and return back to the community, and increase the duration to be deleted from the NPL.

We do not find biases associated with either income, race or education, or the fraction of children. However, we find that the fraction of residents in the neighborhood which is unemployed is a large negative predictor for the duration to deletion, as is the fraction of college educated individuals. Contaminated sites in areas suffering from high unemployment are thus less likely to be returned to general use and may linger on contaminated for quite some time. This reflects the possibility that in already economically depressed areas re-purposing a past Superfund site is not easily accomplished.

The baseline hazard is estimated as a flexible partial linear function in all models but is not reported in the tables due to space limitations. In all models for LC durations we have found the baseline hazard to be monotonically increasing which is consistent with the cleanup following a well-defined process driven by engineering milestones. The baseline hazard for the CD durations however is estimated to be non-monotonic reflecting the fact that after the construction is completed the site undergoes regular but not continuous reviews to determine progress and whether it can be returned to general use. In Section 5.3 we discuss how the estimated unobserved site specific heterogeneity can be interpreted and what insights we can gain from it.

4.2. **Time of Listing**

One important consideration is the fact that the timing of the discovery of Superfund sites is not random. It is thus possible that Superfund sites may spuriously correlate to neighborhood characteristics in virtue of the time when they were listed unless we also control for the year of listing. In Table 5 we present estimation results from models for the two durations of interest that also control for the year of listing.

We find that this virtually does not change the impact of the engineering characteristics of the site such as the cost, contamination type, and contaminated media. We do see, however, some changes in the estimated effect of the demographic features of the neighborhood. When considering the LC duration, we continue to find that neighborhoods with a larger proportion of the population over 65 are cleaned up faster but the relationship to income becomes statistically insignificant. For the CD duration, we continue to find that sites located in areas with high unemployment take longer to be released for general use. We now also find a small negative impact of income.

It is rather surprising that in the above specifications the relationship between income and the two durations of interest is sensitive to the inclusion of the controls for the time of listing. This may indicate that the relationship itself is time varying and requires additional model specifications.

In the history of Superfund there are two distinct periods in the development of the program itself that need to be considered. They are separated by a very important milestone in the development of the Superfund program, Executive Order 12898, “Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations”, signed by President Bill Clinton in February 1994, which directed the attention of federal agencies to issues of environmental equity. In particular it explicitly focuses on the problems faced by low income and minority populations living near a Superfund site.

We explore the effect of the 1994 policy change by interacting an indicator variable capturing the period of listing 1994-2010 with all the demographic variables used in the model (in addition to controlling for the time of listing). If the Executive Order did not change the prioritization of cleanup procedures, we would not expect the interaction terms to be statistically significant. We present the results for the two durations of interest LC and CD in Table 5. For the LC duration, neighborhoods with a high proportion of residents over 65 continue to be cleaned up faster overall, but now it is also the case that sites located in low income areas and areas with high unemployment are cleaned up faster after 1994 than before that year. These large negative coefficients for median income and unemployment indicate that after 1994 the prioritization of resources was effectively directed towards speeding up the cleanups in economically depressed neighborhoods. It is interesting to note that areas with highly educated residents also experience a faster cleanup after 1994. The 1994 policy change also included provisions for greater transparency and community involvement, which seems to be reflected in the faster cleanup durations.

In contrast, the results for the CD duration do not change much with the inclusion of the interaction between the demographics and the post-1994 period, indicating that the policy change had a much smaller impact on the process that leads to a site being deleted from the NPL list. We continue to find that the primary demographic driver is whether a site is located in an economically depressed neighborhood.

Another important feature of the Superfund NPL listing timeline is the distinction between the first listing wave in 1983 and sites that were listed after that year. The initial Superfund site discovery process started already in 1980 but the discovered sites were only listed upon the official launch of the cleanup program in 1983. Beider (1994) interviews site managers who argue that the sites that were initially listed on the NPL were quintessentially different than sites listed in later years and presented a number of technical challenges that had to be overcome which affected the cleanup duration. We therefore split the sample into sites that were listed in 1983 and sites that were listed after that year. We estimate separate models for each split sample for both the LC and CD durations. The estimated coefficients are presented in Tables 6 and 7.

First, consider the results for the LC duration for sites listed in the first wave in 1983. It is particularly notable that the nature of contamination does not appear to drive the durations at all. The only exception consists of sites with contaminated sediment which take longer to clean up. At the same time, the impact of the demographic variables is large and significant. Sites with a large share of urban and black population take much longer to be cleaned up while sites with a highly educated population are cleaned up faster. In contrast, when we consider the sites listed after 1983, it appears that their cleanup duration is driven largely by costs and the nature of the contamination and not by the demographic characteristics. Sites located in neighborhoods with a larger share of the population over 65 are cleaned up faster (although the coefficient is not significant in these specifications). If we now consider the CD duration, we find that for both sites listed before and after 1983, the single largest determinant of the duration is the economic health of the neighborhood as measured by the fraction of the population which is unemployed. For sites listed more recently, the fraction of the population under 18 also seems to be a significant driver for speeding up the release of the site for general use. In both cases contaminated groundwater is a major delay factor.

4.3. Economic Significance

The models estimated above reveal that the factors identified to be statistically significant in driving the durations between the different milestones in the cleanup process are also economically very significant. We use the methodology described in Section 3.5 to quantify the economic significance of a discrete change in a variable of interest and determine what the implied counterfactual change in the expected cleanup duration is. A one standard deviation increase in the expected cost of a cleanup increases the LC cleanup duration by 4.8 years. Similarly the effects of the contaminants and contaminated media are also very significant. The presence of metal increases the LC duration by 1.4 years while the presence of radioactive substances increases the duration by 5.1 years. The CD duration is somewhat less determined by cost and contaminants. A one standard deviation increase in the expected cost increases the CD duration by 1.8 years. The contaminated media is however much more important. Contaminated groundwater increases the CD duration by an average of 5.3 years.

The impact of the demographics is also substantial. An increase in one standard deviation in the fraction of the population over 65 reduces the LC duration by approximately 8 months. In contrast a one standard deviation in the fraction of the population that is unemployed delays deletion from the NPL list by an average of 3.6 years.

After 1994 we see that sites located in neighborhoods which are one standard deviation poorer reach the construction complete milestone 1.2 years faster. Similarly, sites in neighborhoods with higher unemployment also have an LC duration which is 2-3 months shorter. In contrast we measure that sites listed in 1983 reached the first cleanup milestone 5.7 years sooner in more educated neighborhoods but 2.5 years later in areas with a higher black population. This confirms that the economic significance of the observed demographic discrimination during the initial phase of the Superfund program was quite substantial.

5. The Role of Bargaining Power

To the extent that we found that cleanup durations are a function of community characteristics, it is important to assess whether the estimated effects are a result of policy bias in terms of the implementation of cleanup activities or whether they result from the differential use of bargaining power by the parties involved in the cleanup (including the community). The first possibility would be an indicator of direct discrimination based on neighborhood demographics, while the second might reflect the extent to which different parties are involved in the process itself while the degree of involvement may correlate with the demographic characteristics. From an econometric perspective, if the demographic variables are really capturing the degree to which the parties influence the cleanup process, we would expect that once we control for proxies describing the involvement of the different parties, the effect of the demographics will diminish.

Below we consider two measures of involvement. One characterizes the litigation process associated with the cleanup, and the other measures the extent to which the communities were actively involved in deciding the course of the cleanup.

5.1. **Litigation**

The EPA searches for the Principal Responsible Parties (PRP) associated with a Superfund site as a part of the litigation process. Following a letter of determination these parties are asked to contribute financially to the cleanup. It is important to note that in many cases no such responsible parties can be found. This is generally because the associated entities no longer exist, such as companies that dumped hazardous waste but have since been dissolved. If the parties refuse to pay, legal action will be initiated.

While the EPA list of PRPs is available, it is not possible to find out detailed information about these companies in a comprehensive fashion. Most of the parties are quite small and no longer exist. Thus, they are not tracked by databases such as Bloomberg or Compustat. With these limitations in mind we create indicator variables for the case where no PRP exists for a site (PRP 0), where the number of parties is between 2-10 (PRP 2-10), and the case where the number of parties is greater than 10 (PRP 10+). These provide a rough approximation of the liability share of each party which will then impact the subsequent litigation and potentially cleanup duration.

In Table 8 we show the coefficient estimates for both the baseline model and the model with year of listing indicators for both the LC and CD durations where we add the PRP indicators. We find that sites with more than 10 PRPs experience faster construction completion times but that the number of PRPs does not influence the time it takes to return the site to general use. Since litigation happens at the beginning of the cleanup process, it makes sense for the litigation process to only affect the LC but not the CD durations.

At first glance it may seem counterintuitive, that a larger number PRPs is associated with shorter cleanup durations. This is consistent with the existing literature on Superfund litigation though, which suggests that the existence of multiple parties does improve the odds of settlement thereby reducing the length of the litigation process and reducing the LC duration (Rausser and Simon 1998, Sigman 1998, Chang and Sigman 2000). The intuition is that it is easier to obtain settlements from litigation with many small parties than one large corporation which can sustain a prolonged court battle. When sites have a small number of PRPs, it usually indicates that the site is owned by a large corporation. In such a case, as earlier studies have shown, the large corporation has an incentive to minimize its liability and require lengthy reviews, thereby delaying the cleanup process. Furthermore, the presence of many PRPs can also be associated with mostly local entities who may have a more direct concern or benefit from the the cleanup. The impact of the joint liability framework is also economically significant. Sites where the number of PRPs is larger than 10 complete the LC duration an average of 2 years earlier.

Concerning our main hypothesis, we seek to assess whether the observed demographic biases reflect policy biases or are driven by the extent to which neighborhoods with different demographic characteristics are also host to different types of businesses. Since the litigation process involves the PRPs operating in that community, delays due to the litigation process may be falsely attributed to neighborhood characteristics. Table 8 however reveals that this is not the case. The coefficients on the demographic variables do not change much with the addition of the PRP variables.

5.2. **Community Involvement**

While we do not have a direct measure of the extent to which a community is concerned about the timing and nature of the cleanup of a local Superfund site, we do attempt to proxy for community involvement in two different ways. First, we investigate whether the fraction of home ownership in the community impact the cleanup durations. We report the estimated coefficients in Table 9. While home ownership does not have a significant effect on the LC duration it does increase the probability that a site is deleted from the NPL list substantially. A one standard deviation increase in the proportion of homeowners reduces the CD duration by almost 4 years. We note moreover that adding home ownership to the model leads to a small decline of the effect of the percent of the population over 65 on the CD duration, but since it is highly correlated with the percent of the population which is unemployed it removes the statistical significance of the latter variable in the CD specification. This makes it difficult to interpret home ownership as a proxy for community involvement. While the results appear to suggest that homeowners are somewhat more likely to be involved in the cleanup process, it is also likely that the variable captures an aspect of the economic vibrancy of a community.

Second, we evaluate the extent to which the community was involved in the cleanup decision process as recorded by the EPA. This involvement can happen at any point in the process but does require coordination with the EPA site manager. Community involvement can take many forms of dialogue between the EPA and the public such as public meetings. The data does not record precise details on the process of community involvement, but it does report whether community relations activities were conducted to address concerns raised by the local community.

Using the available data, we construct a site specific indicator which records whether the community was involved in the cleanup process. Since Executive Order 12898 placed a much heavier emphasis on community involvement as part of its requirement to promote Environmental Justice, we also create an indicator variable which captures whether community relations activities were performed for sites listed after 1994.

In Table 10, we report results for both the baseline model and the model with year of listing indicators for both the LC and CD durations. We add the above indicators for community involvement and find that community involvement is a significant predictor of shorter LC durations, but not for CD durations in the models which account for the year of listing. Moreover, the magnitude of this effect is several times larger after 1994. This reflects the extent to which community involvement was made a policy priority after 1994. At the same time, for the LC duration model which controls for time of listing, we see that adding controls for community involvement removes the statistical significance of the demographic variables. The coefficient on the fraction of the population over 65 is reduced from 3.234 to 1.424 and becomes statistically insignificant. We do not find a corresponding effect for community involvement on the CD duration. The impact of community involvement is economically significant. Before 1994 we estimate that sites with active community involvement completed the LC duration on average one year earlier than sites without community involvement. After 1994 sites with community involvement activities reached the first cleanup milestone on average of 5.4 years sooner.

This indicates that community involvement plays an important role in explaining the heterogeneity between cleanup durations, even after accounting for technical factors related to the nature and extent of the contamination. It is difficult to interpret this finding causally, however, since community activities are often initiated by the EPA site manager. Thus, while it is certainly probable that communities with a population over 65 are more likely to be engaged in the cleanup process and participate in community events, we cannot exclude the possibility that at least some neighborhoods were discriminated against by not engaging the local community in the cleanup process. The analysis seems to confirm this view by finding a much larger impact of community involvement after 1994, when Environmental Justice considerations prioritized community involvement in the cleanup process.

5.3. Unobserved Site Heterogeneity

Figure 2 shows the nonparametric estimate of the unobserved site heterogeneity estimated from each of the baseline models corresponding to the two durations of interest. The density estimate indicates that the distribution of heterogeneity can be characterized by two modes and a thick right tail. Thus, a small number of sites corresponding to heterogeneity estimates close to zero suffer from conditions which slow down the clean up process. At the other extreme, there is a substantial number of sites that benefit from additional unobserved factors that speed up the cleanup process.

The estimated unobserved individual heterogeneity of Superfund sites can be interpreted as a factor which also contributes to the variation in the cleanup or deletion duration but is not included among the observable explanatory variables. The heterogeneity term thus acts as another explanatory variable in itself, albeit not directly measured but rather inferred indirectly from the model. The distribution of heterogeneity across all sites is normalized to have mean one, reflecting the multiplicative way in which it enters the hazard model parameterization. Its influence is exhibited as deviations beyond the mean effects captured by the measured observables and the baseline hazard parameters. Heterogeneity is thus essentially estimated as explaining the deviations of durations from the mean model prediction once the effect of the observables has been accounted for. We do not constrain the distribution of heterogeneity to any specific parametric shape, but rather endow it with a flexible nonparametric model in order to mitigate any potential model misspecification biases. At the same time, under the Bayesian hierarchical model framework, the distribution of unobserved heterogeneity is allowed to be correlated with the observed explanatory variables. An analysis of this correlation pattern may indicate the source of heterogeneity.

In a post-estimation analysis, we investigate the extent to which the estimated heterogeneity at the individual site level correlates with the site and neighborhood characteristics by regressing individual heterogeneity on the full set of covariates for each type of duration. Statistically significant partial correlation of heterogeneity was detected for some demographic characteristics of the neighborhoods with Superfund sites for the completion to deletion duration, namely income (negative), higher education (positive), and fraction of urban population (positive). This suggests that the influence of the unobserved individual component on faster deletion duration decreases with higher income but increases with education and urbanization.

It is difficult to interpret the exact meaning of the unobserved individual component. Nonetheless, since virtually no heterogeneity correlation was detected for the site physical characteristics we can conclude that the influence of any unobservables beyond the mean effect captured in the main model rests either with the neighborhood characteristics (as opposed to the site attributes) or other factors orthogonal to the variables included in the model.

States are also involved to some degree in the cleanup process and thus one possibility is that the unobserved heterogeneity captures funding or political economy differences across States. However, we could not detect any statistically significant differences between the State level heterogeneity averages across States. Any State mean differences in terms of the observables (such as income or fraction of urban population) are controlled for at the individual site level and it appears that there is no residual spatial pattern of unobserved differences on the aggregate level.

6. **Conclusion**

This paper introduces a more nuanced analysis of Environmental Justice in Superfund cleanups than has previously been available. Given the inherent demographic bias resulting from the geographic location decisions made by firms producing hazardous waste, we focus on the duration of Superfund cleanups which is subject to decisions made by the various parties involved in the cleanup process.

Our identification assumption relies on the observation that conditional on a large number of observable site characteristics, a rational cleanup process subject to cost-benefit analysis will depend only on the site characteristics and not on the demographic composition of the neighborhood. We use a state of the art econometric model to further account for the presence of unobserved site heterogeneity.

The empirical results strongly suggest that the nature of demographic biases changed over time. In particular we find that the cleanup of Superfund sites listed in the initial phase of the program in the early 1980s suffered from a number of biases against sites located in black, urban neighborhoods but in favor of sites located in areas with a highly educated population. These biases appear to diminish over time however, largely following the 1994 Executive Order which formally establishes Environmental Justice as a policy concern. After 1994 we see in fact a prioritization of cleanups in economically disadvantaged neighborhoods. Furthermore, some of these biases may have manifested themselves through the extent to which the community was involved with the cleanup process. We do not find the associated litigation process to be an impediment to Superfund cleanups. The return of a site to general use remains slow and driven by the overall economic health of the community. This suggests that additional resources ought to be made available to assist with the process of deleting Superfund sites from the NPL list in underprivileged areas.

While we believe that, in general faster, cleanups are beneficial to the communities where Superfund sites it is important to note that based on the analysis in this paper we do not have the ability to make concrete social welfare statements. Although we don't have any data or evidence to this effect, we cannot exclude the possibility that longer durations may in fact be associated with higher quality cleanups, or reflect unobserved underlying preferences or sensitivity to environmental damage of the communities involved.

References

Anderton, D., J. Oakes, and K. Egan (1997): "Environmental equity in Supefund: Demographics of the discovery and prioritization of abandoned toxic sites," *Evaluation Review*, 21(3), 3–26.

Beider, P. (1994): "Analyzing the duration of cleanup at sites on Superfund's National Priorities List," CBO Memorandum, Congressional Budget Office.

Burda, M., M. Harding, and J. Hausman (2012): "A Bayesian Semi-Parametric Competing Risk Model with Unobserved Heterogeneity," mimeo.

Chang, H., and H. Sigman (2000): "Incentives to settle under joint and deveral liability: An empirical analysis of Superfund litigation," *Journal of Legal Studies*, 29(1), 205–236.

Currie, J., M. Greenstone, and E. Moretti (2011): "Superfund cleanups and infant health," *American Economic Review: Papers and Proceedings*, 101(3), 435–441.

Gayer, T., J. T. Hamilton, and W. K. Viscusi (2000): "Private values of risk tradeoffs at Superfund sites: Housing market evidence on learning about risk," *Review of Economics and Statistics*, 82(3), 439–451.

Greenstone, M., and J. Gallagher (2008): "Does hazardous waste matter? Evidence from the housing market and the Superfund program," *Quarterly Journal of Economics*, 123, 951–1003.

Hamilton, J. T. (1993): "Politics and social costs: Estimating the impact of collective action on hazardous waste facilities," *Rand Journal of Economics*, 24(1), 101–125.

Hamilton, J. T., and W. K. Viscusi (1999): *Calculating risks? The spatial and political dimensions of hazardous waste policy*. MIT Press.

Han, A., and J. A. Hausman (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics*, 5(1), 1–28.

Hougaard, P. (2000): *Analysis of multivariate survival data*. Springer, New York.

Judy, M., and K. Probst (2009): "Superfund at 30," *Vermont Journal of Environmental Law*, 11, 191–247.

Rausser, G., and L. Simon (1998): "Information asymmetries, uncertainties, and cleanup delays at Superfund sites," *Journal of Environmental Economics and Management*, 35, 48–68.

Sigman, H. (1998): "Liability funding and Superfund clean-up remedies," *Journal of Environmental Economics and Management*, 35, 205–224.

——— (2001): "The pace of progress at Superfund Sites: Policy goals and interest group influence," *Journal of Law and Economics*, 44(1), 315–343.

Sigman, H., and S. Stafford (2010): "Management of Hazardous Waste and Contaminated Land," mimeo.

Smith, C. (2009): "Economic deprivation and racial segregation: Comparing Superfund sites in Portland, Oregon and Detroit, Michigan," *Social Science Research*, 38, 681–692.

Stretsky, P., and M. Hogan (1998): "Environmental Justice: An analysis of Superfund sites in Florida," *Social Problems*, 45(2), 268–287.

United Church of Christ (UCC) (1987): *Toxic wastes and race in the United States: A national report on the racial and socioeconomic characteristics of communities with hazardous waste sites*. Commission for Racial Justice.

VISCUSI, W. K., AND J. T. HAMILTON (1999): “Are risk regulators rational? Evidence from hazardous waste cleanup decisions,” *American Economic Review*, 89(4), 1010–1027.

WOLVERTON, A. (2009): “Effects of socio-economic and input related factors on polluting plants’ location decisions,” *B.E. Journal of Economic Analysis and Policy*, 9(1), 1–30.

Table 1: Summary Statistics

	1980		1990		2000	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
hrs	41.044	9.357	44.306	9.744	47.915	7.931
Cost ($m)	15.840	8.772	8.624	9.896	11.909	15.560
Acids	0.490	0.500	0.365	0.482	0.214	0.415
Dioxins Dibenzofurans	0.133	0.339	0.150	0.358	0.119	0.327
Inorganics	0.338	0.473	0.296	0.457	0.071	0.260
Metals	0.775	0.417	0.772	0.420	0.738	0.445
PAH	0.555	0.497	0.520	0.500	0.333	0.477
PCBs	0.320	0.466	0.247	0.432	0.095	0.297
Pesticides	0.299	0.458	0.308	0.462	0.214	0.415
Radioactive	0.040	0.196	0.056	0.232		
VOC	0.798	0.401	0.796	0.403	0.571	0.500
Other Contaminants	0.170	0.376	0.154	0.362	0.142	0.354
Debris	0.188	0.391	0.093	0.291		
Groundwater	0.863	0.344	0.873	0.332	0.714	0.457
Sediment	0.320	0.466	0.329	0.470	0.214	0.415
Surface Water	0.249	0.432	0.247	0.432	0.166	0.377
Soil	0.797	0.402	0.768	0.422	0.809	0.397
Waste	0.232	0.422	0.105	0.308	0.095	0.297
Other Contaminated Media	0.130	0.337	0.109	0.313	0.190	0.397
N	774		246		42	
Household Median Income	36,767	10,134	24,217	8,837	19,306	7,020
Fraction of Unemployed	0.041	0.019	0.039	0.019	0.035	0.023
Fraction of Bachelor plus	0.051	0.030	0.112	0.074	0.139	0.089
Fraction of Black	0.079	0.149	0.087	0.153	0.094	0.158
Fraction of Urban	0.456	0.457	0.656	0.367	0.722	0.334
Fraction Age 0-17	0.292	0.051	0.260	0.047	0.250	0.046
Fraction Age 65 plus	0.104	0.045	0.122	0.047	0.144	0.052
N	1062		1062		1062	

Table 2: Demographics during the cleanup period.

	Change over time				Correlation	
	All	$LC \leq 10$	$10 < LC \leq 20$	$LC > 20$	All	Listed after 1983
ln(income)	-0.484	-0.414	-0.463	-0.595	-0.346	-0.324
Fraction of Unemployed	-0.003	-0.001	-0.005	-0.003	-0.029	-0.012
Fraction of Bachelor plus	0.065	0.053	0.059	0.088	0.231	0.082
Fraction of Black	0.007	0.009	0.005	0.011	0.033	0.022
Fraction of Urban	0.188	0.159	0.187	0.220	0.084	0.091
Fraction Age 0-17	-0.031	-0.030	-0.030	-0.033	-0.066	-0.025
Fraction Age 65 plus	0.027	0.016	0.028	0.036	0.254	0.216

LC denotes list to construction completion duration ($N = 1,062$, uncensored = 787, censored = 275).

Table 3: Models without Site Heterogeneity

Model Type	LC		CD		LC		CD	
Variable	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
hrs					-0.013**	0.004	-0.002	0.009
ln(cost)					-0.198**	0.021	-0.213**	0.039
Cost zero indicator					-2.515**	0.369	-1.487**	0.596
Acids					0.146*	0.084	-0.126	0.172
Dioxins Dibenzofurans					-0.119	0.110	0.247	0.246
Inorganics					0.110	0.083	0.171	0.161
Metals					-0.234	0.099	0.427**	0.197
PAH					0.086	0.092	0.088	0.166
PCBs					-0.216**	0.093	0.098	0.176
Pesticides					-0.154	0.094	-0.328	0.208
Radioactive					-0.663**	0.204		
VOC					-0.171	0.109	-0.451**	0.191
Other Contaminants					-0.399**	0.113	0.251	0.204
Debris					0.253**	0.096	-0.406*	0.213
Groundwater					-0.098	0.121	-1.185**	0.194
Sediment					-0.274**	0.089	-0.035	0.174
Surface Water					-0.104	0.094	0.183	0.192
Soil					-0.245**	0.087	0.070	0.189
Waste					0.171*	0.087	0.137	0.197
Other contaminated media					-0.129	0.117	-0.378*	0.238
ln(income)	-0.652**	0.126	-0.250*	0.143	-0.517**	0.105	0.403**	0.186
Fraction of Unemployed	-3.996*	2.094	-6.137	3.854	-1.597	2.076	-6.026	3.885
Fraction of Bachelor+	-0.924	1.029	-3.639*	2.044	0.468	0.995	-3.633*	2.051
Fraction of Black	-0.477*	0.277	0.656	0.463	-0.174	0.288	0.978*	0.480
Fraction of Urban	-0.285**	0.085	-0.060	0.175	-0.156	0.098	0.181	0 .17
Fraction Age 0-17	-1.457	1.004	2.439	1.529	-0.569	1.000	2.347	1.666
Fraction Age 65 plus	-0.727	0.939	-0.728	1.421	0.412	0.929	1.930	1.567

LC denotes list to construction completion duration ($N = 1,062$, uncensored = 787, censored = 275).
CD denotes construction completion to deletion duration ($N = 787$, uncensored = 205, censored = 582).

Table 4: Base Model with Site Heterogeneity

Model Type	Parametric LC		Non-parametric LC		Parametric CD		Non-parametric CD	
Variable	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
hrs	-0.012**	0.005	-0.011**	0.004	-0.007	0.011	-0.007	0.009
ln(cost)	-0.191**	0.028	-0.180**	0.025	-0.368**	0.057	-0.293**	0.049
Cost zero indicator	-2.121**	0.470	-2.068**	0.431	-2.929**	0.863	-2.434**	0.746
Acids	0.131	0.110	0.106	0.100	-0.040	0.279	-0.117	0.184
Dioxins Dibenzofurans	-0.117	0.151	-0.092	0.125	0.316	0.364	0.298	0.294
Inorganics	0.094	0.106	0.086	0.094	0.256	0.214	0.193	0.177
Metals	-0.226*	0.122	-0.212**	0.107	0.489**	0.251	0.424**	0.202
PAH	0.016	0.119	0.035	0.109	0.119	0.247	0.111	0.196
PCBs	-0.289**	0.117	-0.237**	0.101	0.191	0.265	0.163	0.192
Pesticides	-0.169	0.120	-0.154	0.102	-0.485	0.316	-0.381*	0.230
Radioactive	-0.914**	0.255	-0.710**	0.216				
VOC	-0.182	0.139	-0.164	0.123	-0.652**	0.276	-0.515**	0.208
Other Contaminants	-0.394**	0.135	-0.371**	0.112	0.413	0.348	0.262	0.246
Debris	0.278*	0.131	0.252**	0.114	-0.631**	0.302	-0.445**	0.214
Groundwater	-0.057	0.150	-0.068	0.134	-1.918**	0.311	-1.371**	0.221
Sediment	-0.363**	0.117	-0.319**	0.102	-0.030	0.294	-0.054	0.208
Surface Water	-0.072	0.123	-0.070	0.110	0.306	0.277	0.183	0.213
Soil	-0.234*	0.122	-0.215**	0.101	-0.058	0.287	0.042	0.207
Waste	0.198	0.120	0.189*	0.107	0.309	0.276	0.197	0.200
Other contaminated media	-0.181	0.139	-0.150	0.122	-0.460	0.360	-0.362	0.298
ln(income)	-0.136	0.144	-0.212**	0.102	0.551**	0.269	0.140	0.215
Fraction of Unemployed	0.770	2.454	0.762	2.274	-7.405	5.554	-7.624*	4.462
Fraction of Bachelor+	2.018*	0.886	1.581	1.001	-4.636	2.956	-4.476*	2.422
Fraction of Black	0.105	0.351	0.023	0.310	1.256*	0.740	0.917	0.548
Fraction of Urban	-0.051	0.118	-0.080	0.101	0.242	0.288	0.129	0.200
Fraction Age 0-17	2.406*	1.312	1.502	0.995	4.132*	2.573	1.253	1.931
Fraction Age 65 plus	4.253**	1.390	2.930**	0.916	2.801	2.532	-0.031	1.869

LC denotes list to construction completion duration ($N = 1,062$, uncensored = 787, censored = 275).

CD denotes construction completion to deletion duration ($N = 787$, uncensored = 205, censored = 582).

Table 5: Base Model with List Year Dummies

Model Type	LC		CD		LC		CD	
Variable	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
hrs	-0.005	0.005	-0.001	0.009	-0.005	0.005	-0.003	0.009
ln(cost)	-0.218**	0.030	-0.265**	0.053	-0.222**	0.030	-0.288**	0.051
Cost zero indicator	-2.438**	0.497	-1.916**	0.796	-2.294**	0.494	-2.175**	0.805
Acids	0.093	0.104	-0.165	0.190	0.100	0.107	-0.152	0.202
Dioxins Dibenzofurans	-0.158	0.139	0.335	0.253	-0.139	0.141	0.347	0.272
Inorganics	0.054	0.103	0.202	0.179	0.057	0.103	0.212	0.187
Metals	-0.239**	0.118	0.451**	0.209	-0.229**	0.115	0.470**	0.214
PAH	0.011	0.119	0.112	0.196	-0.005	0.113	0.112	0.208
PCBs	-0.220**	0.111	0.068	0.204	-0.184*	0.109	0.089	0.213
Pesticides	-0.155	0.110	-0.301	0.218	-0.161	0.112	-0.320	0.232
Radioactive	-0.915**	0.241			-0.891**	0.241		
VOC	-0.235	0.140	-0.444**	0.213	-0.201	0.134	-0.461**	0.229
Other Contaminants	-0.397**	0.135	0.222	0.232	-0.396**	0.134	0.241	0.245
Debris	0.249**	0.117	-0.450**	0.223	0.274*	0.119	-0.479**	0.237
Groundwater	-0.143	0.153	-1.349**	0.211	-0.169	0.153	-1.415**	0.223
Sediment	-0.342**	0.112	0.017	0.206	-0.355**	0.113	-0.007	0.213
Surface Water	-0.077	0.116	0.210	0.209	-0.099	0.116	0.224	0.218
Soil	-0.211*	0.112	-0.005	0.204	-0.241**	0.114	-0.012	0.214
Waste	0.213*	0.114	0.147	0.193	0.210*	0.114	0.160	0.198
Other contaminated media	-0.202	0.132	-0.413	0.266	-0.199	0.134	-0.411	0.287
ln(income)	0.018	0.160	-0.495*	0.273	0.144	0.154	-0.615**	0.279
Fraction of Unemployed	2.995	2.403	-7.468*	4.360	2.280	2.407	-9.013**	4.334
Fraction of Bachelor+	0.808	1.437	1.121	2.845	-0.065	1.575	0.473	2.807
Fraction of Black	-0.040	0.348	0.729	0.479	0.122	0.360	0.684	0.570
Fraction of Urban	-0.053	0.115	0.295	0.202	-0.083	0.115	0.272	0.208
Fraction Age 0-17	1.507	1.219	3.472	2.230	0.290	1.184	2.979	2.172
Fraction Age 65 plus	3.234**	1.062	0.830	1.962	2.355**	1.066	0.044	1.983
L1984-86	0.253**	0.122	0.011	0.214	0.243**	0.123	0.026	0.203
L1987-89	0.588**	0.141	-0.236	0.223	0.565**	0.134	-0.250	0.217
L1990-92	0.547**	0.200	-1.010**	0.387	0.615**	0.190	-0.994**	0.352
L1993-95	-0.159	0.299	-3.984**	1.122	1.976	4.100	-0.260	0.852
L1996+	0.710**	0.283	-1.409**	0.636	2.373	4.063	0.103	0.785
L94-10×ln(income)					-1.824**	0.890	-0.566	0.364
L94-10×Fraction of Unemployed					3.709**	1.098	-0.407	0.912
L94-10×Fraction of Bachelor+					7.243**	3.723	-0.049	1.029
L94-10×Fraction of Black					-0.421	1.159	0.177	0.943
L94-10×Fraction of Urban					-0.362	0.455	-0.284	0.769
L94-10×Fraction Age 0-17					7.651	5.305	-0.099	1.014
L94-10×Fraction Age 65 plus					7.672	5.906	0.009	0.813

LC denotes list to construction completion duration ($N = 1,062$, uncensored = 787, censored = 275).

CD denotes construction completion to deletion duration ($N = 787$, uncensored = 205, censored = 582).

Table 6: Split Samples, List to Construction Completion Duration

Model Type	List Year 1983		List Years 1984–2010	
Variable	Mean	s.e.	Mean	s.e.
hrs	0.001	0.008	-0.006	0.006
ln(cost)	-0.274**	0.055	-0.191**	0.034
Cost zero indicator	-2.650**	0.983	-2.142**	0.578
Acids	0.082	0.174	0.100	0.125
Dioxins Dibenzofurans	0.025	0.232	-0.190	0.161
Inorganics	-0.111	0.183	0.166	0.114
Metals	-0.307	0.204	-0.217	0.137
PAH	-0.007	0.203	-0.007	0.144
PCBs	-0.086	0.172	-0.290**	0.138
Pesticides	0.145	0.181	-0.258*	0.140
Radioactive	-0.554	0.548	-0.968**	0.279
VOC	0.276	0.231	-0.314**	0.152
Other Contaminants	0.167	0.189	-0.632**	0.172
Debris	-0.120	0.204	0.340	0.144
Groundwater	-0.350	0.230	0.034	0.192
Sediment	-0.632**	0.180	-0.136	0.136
Surface Water	-0.203	0.204	-0.008	0.135
Soil	-0.146	0.194	-0.211	0.143
Waste	-0.141	0.190	0.286**	0.130
Other contaminated media	-0.151	0.228	-0.276*	0.163
ln(income)	0.797	0.296	-0.247	0.187
Fraction of Unemployed	-2.889	4.207	2.624	3.122
Fraction of Bachelor+	13.130**	3.487	1.687	1.353
Fraction of Black	-2.244**	0.657	-0.468	0.388
Fraction of Urban	-0.575**	0.193	0.046	0.137
Fraction Age 0-17	-1.218	1.933	0.424	1.394
Fraction Age 65 plus	2.152	1.804	1.757	1.176
L1987-89			0.329**	0.141
L1990-92			0.129	0.209
L1993-95			-0.533	0.330
L1996+			0.280	0.281

For list year 1983, $N = 294$, uncensored = 233, censored = 61.

For list years 1984–2010, $N = 768$, uncensored = 554, censored = 214.

Table 7: Split Samples, Construction Completion to Deletion Duration

Model Type	List Year 1983		List Years 1984–2010	
Variable	Mean	s.e.	Mean	s.e.
hrs	-0.015	0.015	0.006	0.012
ln(cost)	-0.167	0.104	-0.317**	0.037
Cost zero indicator	-1.385	1.662	-2.365**	0.514
Acids	-0.121	0.350	-0.167	0.253
Dioxins Dibenzofurans	0.603	0.474	-0.002	0.369
Inorganics	-0.094	0.338	0.367	0.221
Metals	0.718**	0.358	0.269	0.253
PAH	-0.043	0.369	0.189	0.245
PCBs	-0.321	0.352	0.382	0.256
Pesticides	-0.886**	0.409	0.015	0.275
VOC	-0.021	0.467	-0.593**	0.260
Other Contaminants	0.482	0.393	-0.088	0.300
Debris	-0.715*	0.424	-0.332	0.249
Groundwater	-1.371**	0 .38	-1.557**	0.270
Sediment	0.331	0.334	-0.247	0.273
Surface Water	0.134	0.387	0.275	0.269
Soil	0.597	0.383	-0.171	0.242
Waste	0.405	0.316	-0.017	0.235
Other contaminated media	0.359	0.381	-0.886**	0.406
ln(income)	-0.175	0.528	-0.386	0.349
Fraction of Unemployed	-16.486**	7.660	-7.334**	3.043
Fraction of Bachelor+	-8.363	7.832	2.446	3.017
Fraction of Black	1.595	1.133	0.696	0.653
Fraction of Urban	0.459	0.403	0.172	0.258
Fraction Age 0-17	-0.996	4.113	5.386**	2.620
Fraction Age 65 plus	-2.007	3.362	2.948	2.225
L1987-89			-0.285	0.221
L1990-92			-1.106**	0.434
L1993-95			-2.907**	1.406

For list year 1983, $N = 233$, uncensored = 68, censored = 165.

For list years 1984–2010, $N = 556$, uncensored = 137, censored = 417.

Table 8: Potentially Responsible Parties (PRP) Variables

Model Type	LC		CD		LC		CD	
Variable	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
hrs	-0.011**	0.004	-0.009	0.009	-0.007	0.005	-0.002	0.009
ln(cost)	-0.190**	0.024	-0.294**	0.047	-0.220**	0.032	-0.248**	0.046
Cost zero indicator	-2.211**	0.391	-2.327**	0.714	-2.425**	0.570	-1.634**	0.706
Acids	0.137	0.096	-0.088	0.195	0.117	0.107	-0.140	0.196
Dioxins Dibenzofurans	-0.087	0.125	0.233	0.285	-0.131	0.137	0.288	0.260
Inorganics	0.100	0.091	0.193	0.187	0.070	0.102	0.197	0.176
Metals	-0.236**	0.104	0.434**	0.218	-0.251**	0.119	0.461**	0.202
PAH	0.047	0.103	0.116	0.200	0.017	0.116	0.091	0.195
PCBs	-0.250**	0.100	0.175	0.215	-0.240**	0.105	0.062	0.204
Pesticides	-0.130	0.102	-0.433*	0.232	-0.134	0.110	-0.322	0.221
Radioactive	-0.664**	0.230			-0.823**	0.244		
VOC	-0.189	0.120	-0.553**	0.210	-0.231	0.133	-0.494**	0.214
Other Contaminants	-0.382**	0.117	0.284	0.233	-0.391**	0.133	0.269	0.229
Debris	0.238**	0.111	-0.451**	0.236	0.228*	0.120	-0.480**	0.232
Groundwater	-0.079	0.129	-1.463**	0.217	-0.146	0.147	-1.364**	0.216
Sediment	-0.322**	0.102	-0.028	0.213	-0.342**	0.109	0.032	0.201
Surface Water	-0.074	0.106	0.206	0.216	-0.080	0.114	0.212	0.208
Soil	-0.229**	0.102	0.022	0.212	-0.246**	0.116	0.032	0.204
Waste	0.173	0.105	0.238	0.204	0.208*	0.110	0.171	0.194
Other contaminated media	-0.172	0.123	-0.389	0.278	-0.227	0.139	-0.404	0.279
ln(income)	-0.347**	0.108	0.158	0.228	-0.158	0.161	-0.161	0.263
Fraction of Unemployed	-0.119	2.192	-7.754*	4.665	2.288	2.319	-6.895	4.284
Fraction of Bachelor+	1.434	0.984	-5.013**	2.556	0.945	1.318	-0.176	2.611
Fraction of Black	-0.078	0.302	0.942*	0.553	-0.164	0.345	0.834	0.528
Fraction of Urban	-0.104	0.098	0.129	0.204	-0.064	0.115	0.297	0.205
Fraction Age 0-17	0.939	0.985	0.935	2.049	1.035	1.148	3.643*	1.969
Fraction Age 65 plus	2.191**	0.916	-0.155	2.096	2.449**	1.057	1.940	1.838
L1984-86					0.253**	0.121	0.014	0.207
L1987-89					0.553**	0.133	-0.227	0.217
L1990-92					0.466**	0.197	-0.793**	0.351
L1993-95					-0.207	0.319	-2.676**	1.362
L1996+					0.661**	0.245	-1.263**	0.648
PRP 0	0.102	0.141	0.271	0.288	0.093	0.158	0.349	0.270
PRP 2–10	0.248	0.152	-0.049	0.302	0.246	0.163	0.000	0.286
PRP 10+	0.328**	0.153	0.233	0.323	0.340**	0.169	0.278	0.300

LC denotes list to construction completion duration ($N = 1,062$, uncensored = 787, censored = 275).

CD denotes construction completion to deletion duration ($N = 787$, uncensored = 205, censored = 582).

Table 9: Home Ownership Variables

Model Type	LC		CD	
Variable	Mean	s.e.	Mean	s.e.
hrs	-0.009	1.005	-0.004	0.009
ln(cost)	-0.213**	0.029	-0.251**	0.040
Cost zero indicator	-2.420**	0.497	-1.826**	0.611
Acids	0.118	0.106	-0.117	0.194
Dioxins Dibenzofurans	-0.143	0.136	0.270	0.267
Inorganics	0.051	0.102	0.157	0.172
Metals	-0.186	0.116	0.469**	0.213
PAH	0.017	0.111	0.097	0.193
PCBs	-0.269**	1.110	0.163	0.218
Pesticides	-0.129	0.113	-0.269	0.220
Radioactive	-0.887**	0.255		
VOC	-0.230*	0.130	-0.514**	0.211
Other Contaminants	-0.386**	0.131	0.365	0.235
Debris	0.277**	0.116	-0.446**	0.212
Groundwater	-0.112	0.147	-1.388**	0.213
Sediment	-0.376**	0.110	-0.075	0.198
Surface Water	-0.013	1.116	0.271	0.207
Soil	-0.194*	0.112	-0.012	0.209
Waste	0.278**	0.115	0.212	0.193
Other contaminated media	-0.174	0.132	-0.235	0.273
ln(income)	0.084	0.160	-0.638**	0.288
Fraction of Unemployed	2.049	2.584	-6.365	4.356
Fraction of Bachelor+	1.441	1.285	-0.110	2.601
Fraction of Black	-0.098	0.356	0.776	0.511
Fraction of Urban	-0.070	0.118	0.303	0.205
Fraction Age 0-17	0.743	1.169	0.856	2.048
Fraction Age 65 plus	2.851**	1.093	1.226	1.838
L1984-86	0.107	0.125	0.104	0.209
L1987-89	0.262*	0.132	-0.049	0.219
L1990-92	0.531**	0.195	-0.403	0.333
L1993-95	0.168	0.503	-3.334**	1.464
L1996+	0.129	0.572	-3.263**	1.377
Fraction of homeowners	-0.219	0.371	1.998**	0.695
L(94-10)×Frac homeown	-0.004	0.713	2.460	1.703

LC denotes list to construction completion duration ($N = 1,056$, uncensored = 786, censored = 270).

CD denotes construction completion to deletion duration ($N = 786$, uncensored = 205, censored = 581).

Table 10: Community Involvement Variables

Model Type	LC		CD		LC		CD	
Variable	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
hrs	-0.011**	0.004	-0.009	0.010	-0.007	0.005	-0.002	0.009
ln(cost)	-0.185**	0.025	-0.346**	0.052	-0.205**	0.026	-0.280**	0.051
Cost zero indicator	-2.134**	0.405	-3.094**	0.786	-2.212**	0.462	-2.093**	0.804
Acids	0.121	0.097	-0.118	0.218	0.110	0.102	-0.161	0.196
Dioxins Dibenzofurans	-0.105	0.127	0.310	0.305	-0.166	0.137	0.335	0.272
Inorganics	0.086	0.094	0.203	0.197	0.051	0.102	0.215	0.187
Metals	-0.215**	0.106	0.464**	0.229	-0.220**	0.114	0.470**	0.212
PAH	0.041	0.106	0.097	0.224	0.005	0.113	0.095	0.204
PCBs	-0.237**	0.101	0.165	0.235	-0.232**	0.106	0.082	0.209
Pesticides	-0.146	0.104	-0.435*	0.250	-0.130	0.111	-0.322	0.228
Radioactive	-0.723**	0.226			-0.910**	0.249		
VOC	-0.193	0.122	-0.514**	0.236	-0.214	0.134	-0.434*	0.223
Other Contaminants	-0.380**	0.118	0.292	0.258	-0.393**	0.134	0.244	0.238
Debris	0.254**	0.111	-0.502**	0.258	0.265**	0.119	-0.472**	0.228
Groundwater	-0.060	0.131	-1.514**	0.242	-0.140	0.154	-1.410**	0.223
Sediment	-0.330**	0.104	-0.063	0.236	-0.337**	0.114	0.020	0.209
Surface Water	-0.071	0.109	0.234	0.239	-0.069	0.117	0.242	0.219
Soil	-0.228**	0.105	0.073	0.235	-0.241**	0.112	0.021	0.210
Waste	0.181*	0.106	0.210	0.208	0.209*	0.119	0.156	0.197
Other contaminated media	-0.170	0.126	-0.382	0.311	-0.238*	0.141	-0.365	0.287
ln(income)	-0.246**	0.109	0.224	0.242	-0.318	0.334	-0.926	1.104
Fraction of Unemployed	0.458	2.242	-8.519*	4.782	-0.030	0.160	-0.539*	0.281
Fraction of Bachelor+	1.548	1.012	-5.472	2.680	2.123	1.846	-8.304**	4.340
Fraction of Black	0.020	0.309	1.163*	0.612	0.618	1.407	1.355	2.455
Fraction of Urban	-0.078	0.104	0.128	0.228	-0.062	0.337	0.694	0.564
Fraction Age 0-17	1.456	1.022	0.679	2.172	-0.038	0.112	0.277	0.214
Fraction Age 65 plus	2.711**	0.946	0.039	1.875	1.424	1.235	3.601*	2.089
L1984-86					2.791**	1.086	1.002	1.759
L1987-89					0.209*	0.123	0.016	0.215
L1990-92					0.591**	0.135	-.2634	0.224
L1993-95					0.520**	0.196	-1.074**	0.369
L1996+					-0.000	0.378	-2.650**	1.437
Community	0.111	0.092	-0.222	0.199	0.185*	0.110	-0.203	0.185
L(94-10)×Community					0.878**	0.300	-1.193	0.837

LC denotes list to construction completion duration ($N = 1,062$, uncensored = 787, censored = 275).
CD denotes construction completion to deletion duration ($N = 787$, uncensored = 205, censored = 582).

Figure 1: Distributions of Durations (in Years) , for the different cleanup milestones.

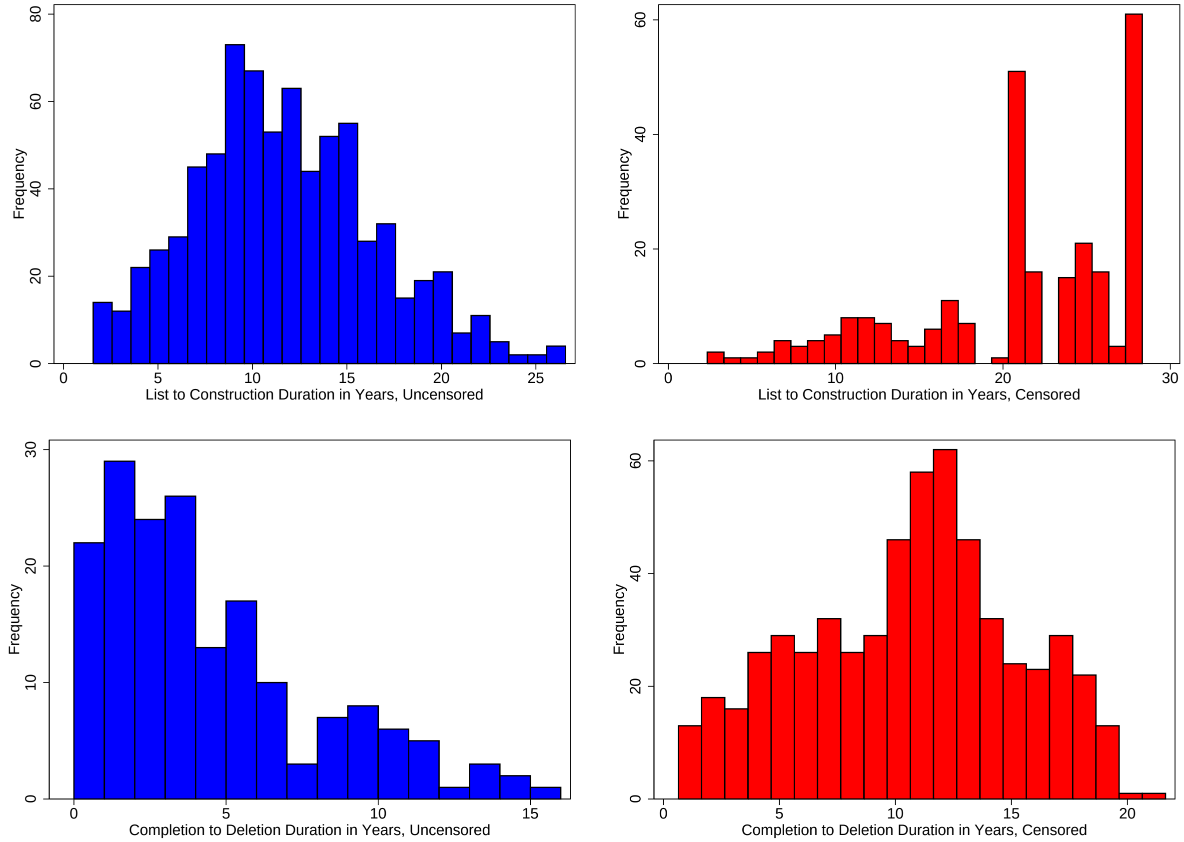
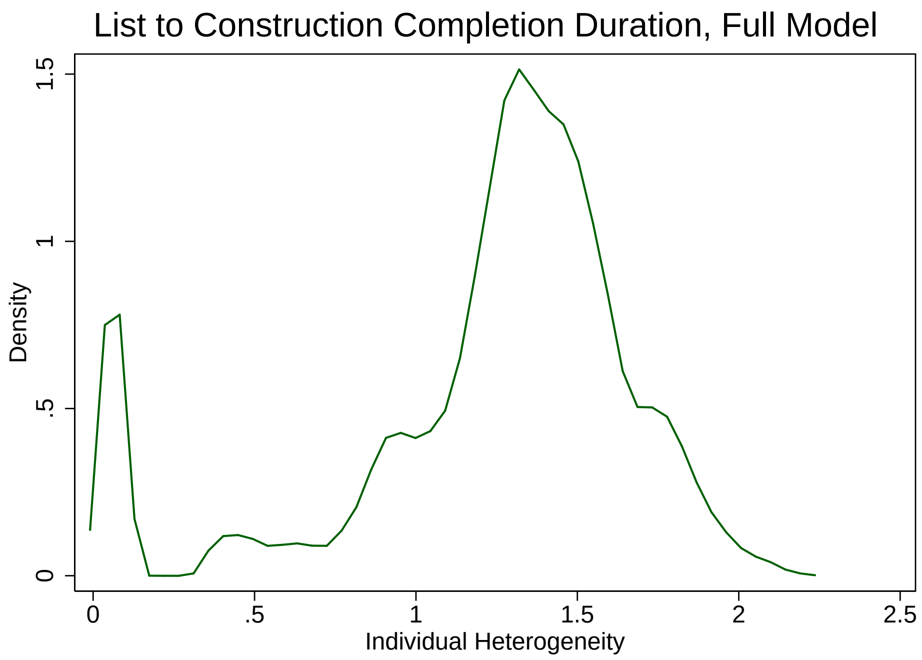
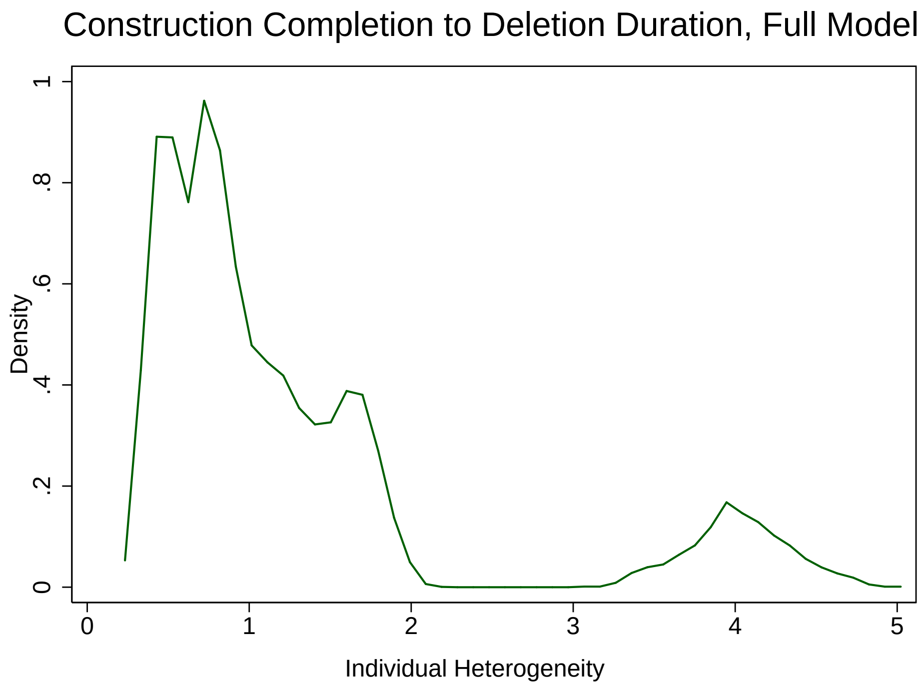


Figure 2: Estimated Density of Individual Heterogeneity


List to Construction Completion Duration, Full Model


Construction Completion to Deletion Duration, Full Model