Learning in a Hedonic Framework: Valuing Brownfield Remediation

Lala Ma

Duke University

August 6, 2013

Camp Resources XX Workshop

What is a Brownfield?

Brownfield - A real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.

- Environmental Protection Agency (EPA)

What is a Brownfield?

Brownfield - A real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.

- Environmental Protection Agency (EPA)

- Examples: gas stations, dry cleaning, factories (shoes, windows, jewelry).

What is a Brownfield?

Brownfield - A real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.

- Environmental Protection Agency (EPA)

- Examples: gas stations, dry cleaning, factories (shoes, windows, jewelry).
- In 2002, the Brownfields Law was enacted to assist organizations in revitalizing brownfields through the provision of grants.

Motivation

Objective: Evaluate the benefits of cleaning up brownfields

- Use hedonic methods - interpret capitalization of brownfield cleanup into housing prices as MWTP for remediation

Motivation

Objective: Evaluate the benefits of cleaning up brownfields

- Use hedonic methods - interpret capitalization of brownfield cleanup into housing prices as MWTP for remediation

Traditional hedonic approach ignores dynamics of housing choices

- Household expectations
- Learning about amenities

Stylized Example

- Suppose brownfield cleaned, but before cleanup, information about hazards released.

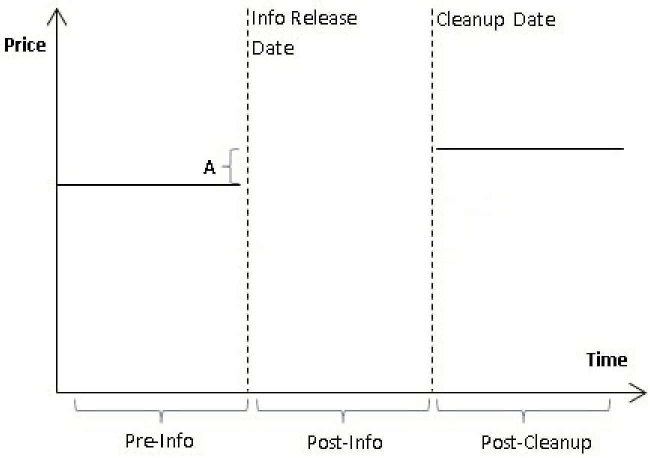
Stylized Example

- Suppose brownfield cleaned, but before cleanup, information about hazards released.
- Static hedonic approach: MWTP = average change in price across cleanup date

Stylized Example

- Suppose brownfield cleaned, but before cleanup, information about hazards released.
- Static hedonic approach: MWTP = average change in price across cleanup date

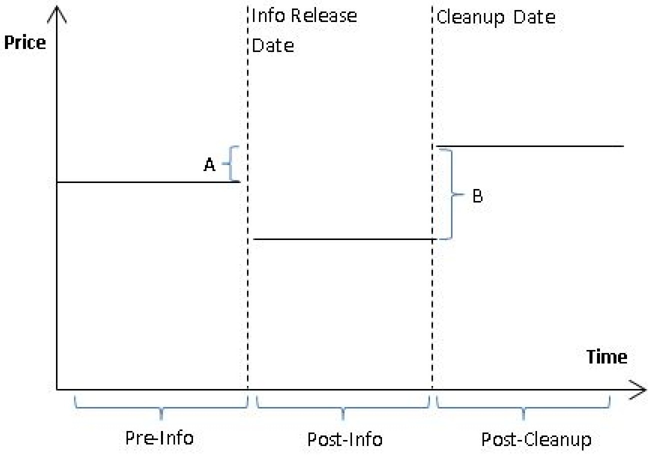
Figure: Households forward-looking, uninformed of contamination until Info Date



Stylized Example

- Suppose brownfield cleaned, but before cleanup, information about hazards released.
- Static hedonic approach: MWTP = average change in price across cleanup date

Figure: Households forward-looking, uninformed of contamination until Info Date



Approach

Use real estate data from Massachusetts, data on brownfield sites, and contamination over time to estimate a model where households

- Learn about brownfield hazards over time through Bayesian updating, and
- Given estimated beliefs, choose residential neighborhoods by maximizing lifetime expected utility.

Research Questions

- Are consumers learning from information released about the brownfields in such a way that may systematically alter the MWTP estimate?
- What is the value of the information that is provided to households?

Contribution

- Add learning into a dynamic, hedonic framework
- Use a newly collected data set on brownfield contamination

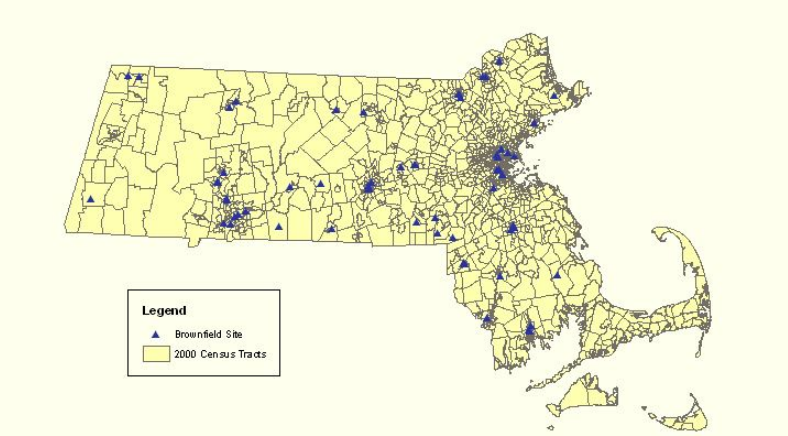
Outline for rest of the talk

- Data for Brownfields
- Model
- Estimation
- Preliminary Results

Outline

Brownfields Sites in Massachusetts

- Cleanup grant applications for 2003 through 2008 (EPA)
- Proposal information: applicant, the proposal score, and start/finish dates of cleanup (if awarded)
- Property information: exact location, property size



Brownfield Contamination over time: Assessments

- To model learning, need to know brownfield contamination over time

Brownfield Contamination over time: Assessments

- To model learning, need to know brownfield contamination over time
- Brownfields are periodically assessed $\longrightarrow$ assessment document

Brownfield Contamination over time: Assessments

- To model learning, need to know brownfield contamination over time
- Brownfields are periodically assessed $\longrightarrow$ assessment document
- Assessment characterize contamination at the time of site investigation
- Each site can have multiple assessments performed

Brownfield Contamination over time: Assessments

Table 1: Brownfield Characteristics

Variable	Obs.	Mean	Median	Std. Dev.	Min.	Max.
Assessments per Site	65	3.43	3	0.98	1	5
Assessment Year	223	2000.08	2002	6.55	1984	2012
Assessment Interval (yrs)	158	4.51	3	3.76	0	18
Contaminant (c_{jt})†	223	2.99	3	2.16	0	10
NRS subscore IV (Institutional)	65	26.2	15	25.58	0	155
NRS subscore V (Environmental)	65	42.2	20	42.82	0	170

† Contaminant (c_{jt}) is the sum of the number of contaminants found in each exposure pathway (soil, groundwater, sediments, air, surface water, or other)

- 65 sites, between 1 and 5 assessments for each site.

Brownfield Contamination over time: Assessments

Table 1: Brownfield Characteristics

Variable	Obs.	Mean	Median	Std. Dev.	Min.	Max.
Assessments per Site	65	3.43	3	0.98	1	5
Assessment Year	223	2000.08	2002	6.55	1984	2012
Assessment Interval (yrs)	158	4.51	3	3.76	0	18
Contaminant (c_{jt})†	223	2.99	3	2.16	0	10
NRS subscore IV (Institutional)	65	26.2	15	25.58	0	155
NRS subscore V (Environmental)	65	42.2	20	42.82	0	170

† Contaminant (c_{jt}) is the sum of the number of contaminants found in each exposure pathway (soil, groundwater, sediments, air, surface water, or other)

- 65 sites, between 1 and 5 assessments for each site.
- c_{jt} = Sum of # contaminants in each contamination pathway

Brownfield Contamination over time: Assessments

Table 1: Brownfield Characteristics

Variable	Obs.	Mean	Median	Std. Dev.	Min.	Max.
Assessments per Site	65	3.43	3	0.98	1	5
Assessment Year	223	2000.08	2002	6.55	1984	2012
Assessment Interval (yrs)	158	4.51	3	3.76	0	18
Contaminant (c_{jt})†	223	2.99	3	2.16	0	10
NRS subscore IV (Institutional)	65	26.2	15	25.58	0	155
NRS subscore V (Environmental)	65	42.2	20	42.82	0	170

† Contaminant (c_{jt}) is the sum of the number of contaminants found in each exposure pathway (soil, groundwater, sediments, air, surface water, or other)

- 65 sites, between 1 and 5 assessments for each site.
- c_{jt} = Sum of # contaminants in each contamination pathway
- NRS scores proxy for potential exposures around site $\longrightarrow$ affects whether site designated as 'contaminated'

Outline

Model: Choice Set

Households choose neighborhood to maximize expected lifetime utility in a finite-horizon framework.

- In each period, households choose whether to move to one of J neighborhoods, or stay in current neighborhood $(J + 1)$
- If a household moves, it incurs a moving cost

Model: Information Structure

- Uncertainty about neighborhood transitions (in the future)

Model: Information Structure

- Uncertainty about neighborhood transitions (in the future)
- In the current period, households are uncertain about brownfield hazard
  - Exists an underlying 'true' brownfield hazard level for each neighborhood.

Model: Information Structure

- Uncertainty about neighborhood transitions (in the future)
- In the current period, households are uncertain about brownfield hazard
  - Exists an underlying 'true' brownfield hazard level for each neighborhood.
- Households do not know the hazard, but can learn about it from published assessments.
  - Assumes once assessment information published, households learn about it.

Model: Assessment Results and Learning

- Published results from assessments serve as noisy signals about brownfield hazards.

Model: Assessment Results and Learning

- Published results from assessments serve as noisy signals about brownfield hazards.
- Assessment results, c_{kt}, for a site k in neighborhood j,

$$c_{kt} = H_k + \lambda \cdot IE_k + e_{kt}$$

  will depend on

  - the unobserved brownfield hazard, H_k
  - the environmental and institutional settings of the site, IE_k
  - noise, e_{kt}, distributed $N(0, \sigma_e)$

Model: Assessment Results and Learning

- The signal for the hazard

$$sig_{kt} = c_{kt} - \lambda \cdot IE_k = H_j + e_{kt}$$

  - If no assessments have been performed for any brownfield in the neighborhood, prior distributed $N(0, \delta)$

Model: Household Preferences

Household i's utility from neighborhood j, located in district r, takes the following form,

$$
\begin{aligned}
u_{ijrt} &= \beta_X X_{jt} + \beta_R R_{jt} + \xi_{jt} + \beta_c E\left(c_{jt} \mid E_t H_j, V_t H_j\right) \times P(noclean_{jt}) \\
&+ \mathbf{1}_{[d_{it} \neq J+1]} \left(\beta_{MC} \cdot MC_{it} + \beta_{PMC} \cdot \mathbf{1}_{[d_{it}^r \neq d_{it-1}^r]}\right) + \beta_d dist(j, d_{it-1}) + \epsilon_{ijt}
\end{aligned}
$$

Model: Household Preferences

Household i's utility from neighborhood j, located in district r, takes the following form,

$$\begin{aligned} u_{ijrt} &= \underline{\beta_X X_{jt} + \beta_R R_{jt} + \xi_{jt} + \beta_c E\left(c_{jt} \mid E_t H_j, V_t H_j\right) \times P(noclean_{jt})} \\ &+ \mathbf{1}_{[d_{it} \neq J+1]} \left(\beta_{MC} \cdot MC_{it} + \beta_{PMC} \cdot \mathbf{1}_{[d_{it}^r \neq d_{it-1}^r]}\right) + \beta_d dist(j, d_{it-1}) + \epsilon_{ijt} \end{aligned}$$

where

- Neighborhood characteristics:
  - X_{jt} (observed), ξ_{jt} (unobserved)
  - Costs of living in the location, R_{jt}
  - Expected contamination, weighted by probability sites will be cleaned

Model: Household Preferences

Household i's utility from neighborhood j, located in district r, takes the following form,

$$
\begin{aligned}
u_{ijrt} &= \beta_X X_{jt} + \beta_R R_{jt} + \xi_{jt} + \beta_c E\left(c_{jt} \mid E_t H_j, V_t H_j\right) \times P(noclean_{jt}) \\
&+ \underline{\mathbf{1}_{[d_{it} \neq J+1]} \left(\beta_{MC} \cdot MC_{it} + \beta_{PMC} \cdot \mathbf{1}_{[d_{it}^r \neq d_{it-1}^r]}\right) + \beta_d dist(j, d_{it-1})} + \epsilon_{ijt}
\end{aligned}
$$

where

- Moving Costs:
  - Financial MC as 6% of value of house last period, MC_{it}
  - Psychological MC if moved to district different from last period, $\mathbf{1}_{[d_{it}^r \neq d_{it-1}^r]}$
  - Distance of the neighborhood from i's previous location, $dist(j, d_{it-1})$

Model: The Household's Problem

Household's problem is to choose a sequence of neighborhood locations (d_{it}) to maximize its expected sum of discounted flow utilities given the state of the world it observes

$$\max_{d_{it}\in\{1,\ldots,J+1\}} \mathbf{E}\left[\sum_{t'=t}^{T} \beta^{t'-t} u(s_{it'}, d_{it'}) + \epsilon_{d_{it'}} \mid \epsilon_{it}, s_{it}, d_{it}\right]$$

- State variables summarized in s_{it},

$$s_{it} = [X_t, R_t, E(H_t), V(H_t), X_t^r, BF_t, \xi_t]$$

Outline

Estimation: 2 stages

1 Get posterior beliefs about brownfield hazards at each point in time

- Expectation-Maximization (EM) Algorithm (Dempster et. al. 1977) Details

Estimation: 2 stages

1. Get posterior beliefs about brownfield hazards at each point in time
   - Expectation-Maximization (EM) Algorithm (Dempster et. al. 1977) Details
2. Estimate dynamic discrete choice problem
   - Hotz and Miller (1993), Arcidiacono and Miller (2011)

Estimation

- Two stage estimation requires that household choices only depend on hazard through the information signals (James, 2012).

Estimation

- Two stage estimation requires that household choices only depend on hazard through the information signals (James, 2012).
- Unobserved neighborhood quality
  - Collapse nbd-time level terms into a mean utility term, θ_{jt} (Berry 1994)
  - Decompose mean utility to recover utility parameters on neighborhood attributes

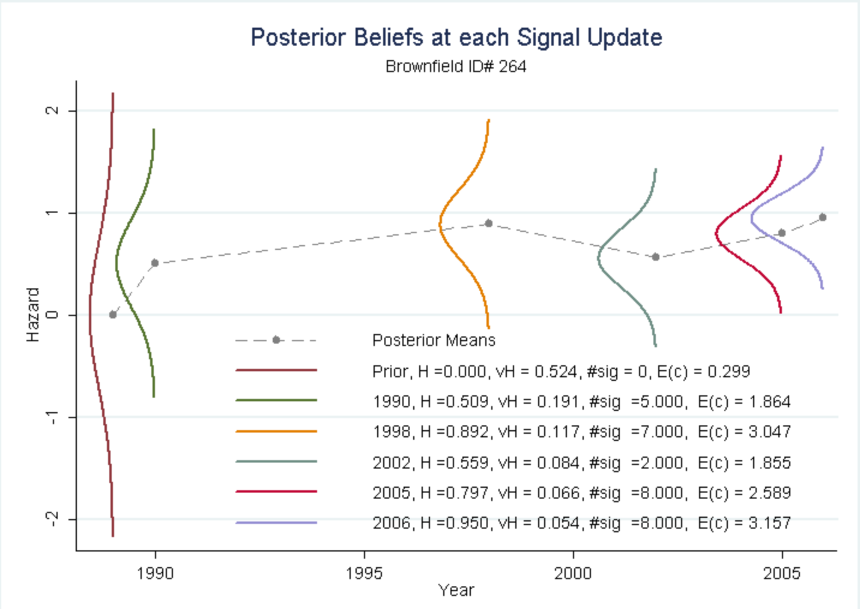
$$u_{ijrt}(s_{it}) = \theta_{jt} + \beta_d dist(j, d_{it-1}) + \mathbf{1}_{[d_{it} \neq J+1]} \left(\beta_{MC} \cdot MC_{it} + \beta_{PMC} \cdot \mathbf{1}_{[d^r_{it} \neq d^r_{it-1}]} \right) + \epsilon_{it}$$

where

$$\theta_{jt} = \beta_X X_{jt} + \beta_R R_{jt} + \beta_c E[c_{jt}] \times Pr(noclean_{jt}) + \xi_{jt}$$

Outline

Preliminary Results: Stage 1



Preliminary Results: Stage 2

Stage 2 Estimates

Discrete Choice Parameters	
β_d (in \$ per km)	\$104.13
β_{PSY} (in \$)	\$11,045.59
β_{MC} (MC in 000's of \$)	-0.2402
θ_{jt}	Not Shown
Mean Utility Decomposition	
Dep. Var.	$\widehat{\theta}_{jt} - \widehat{\beta}_{MC} R_{jt}$
Crime per capita	-28.10
$E(c_{jt}) \times \widehat{Pr}_{noclean_{jt}}$	-0.2038
constant	22.52

District and time period fixed effects included.

Utility Estimates in Dollars	
Decrease in crime per capita	\$60,858.77
Decrease in unit of Contamination	\$421.43

Preliminary Results: Stage 2

Stage 2 Estimates

Discrete Choice Parameters	
β_d (in \$ per km)	\$104.13
β_{PSY} (in \$)	\$11,045.59
β_{MC} (MC in 000's of \$)	-0.2402
θ_{jt}	Not Shown

Mean Utility Decomposition

Dep. Var.	$\widehat{\theta}_{jt} - \widehat{\beta}_{MC} R_{jt}$
Crime per capita	-28.10
$E(c_{jt}) \times \widehat{Pr}_{noclean_{jt}}$	-0.2038
constant	22.52

District and time period fixed effects included.

Utility Estimates in Dollars	
Decrease in crime per capita	\$60,858.77
Decrease in unit of Contamination	\$421.43

Preliminary Results: Comparison

Utility Estimates in Dollars

	Dynamic
Contamination	
No Learning	\$124.24
Learning	\$421.43
Crime (per-person)	\$60,858.77

Value of Information

The DeGroot Measure of value of information:

- Difference in the utility achieved from the optimal choices under pre (τ_0) and post (τ_1) information sets,

$$V_{it}^{I} = \sum_{i} \left[U_i \left(\tau_1, d_{it}(\tau_1) \right) - U_i \left(\tau_1, d_{it}(\tau_0) \right) \right]$$

Value of Information

The DeGroot Measure of value of information:

- Difference in the utility achieved from the optimal choices under pre (τ_0) and post (τ_1) information sets,

$$V_{it}^{I} = \sum_{i} \left[U_i \left(\tau_1, d_{it}(\tau_1) \right) - U_i \left(\tau_1, d_{it}(\tau_0) \right) \right]$$

- Value of an assessment

Value of Information

The DeGroot Measure of value of information:

- Difference in the utility achieved from the optimal choices under pre (τ_0) and post (τ_1) information sets,

$$V_{it}^{I} = \sum_{i} \left[U_i\left(\tau_1, d_{it}(\tau_1)\right) - U_i\left(\tau_1, d_{it}(\tau_0)\right) \right]$$

- Value of an assessment =

$$U(\text{optimal choices given all assessments that occurred}) \\ - U(\text{optimal choices given last assessment not performed})$$

Value of Information

- Do this for site #15

Value of Information per household	\$9,741.15
# of Households near Site #15	3454

'near' = within 3km.

Takeaways

- Learning has a non-trivial effect on MWTP

Takeaways

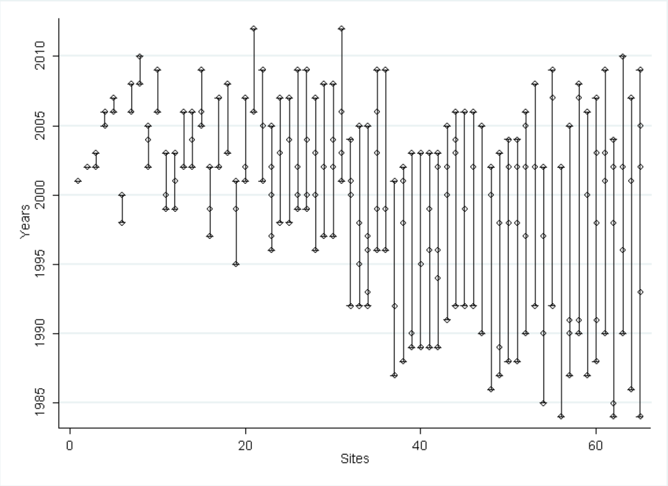
- Learning has a non-trivial effect on MWTP
- Information provision is valuable to households when making housing decisions

Thank you for listening!

I gratefully acknowledge a fellowship from Resources for the Future for the 2013-2014 academic year.

Outline

Interval of Assessments



Model: The Household's Problem

Household's problem is to choose a sequence of neighborhood locations (d_{it}) to maximize its expected sum of discounted flow utilities given the state of the world it observes subject to a budget constraint in each period,

$$\max_{d_{it} \in \{1,\dots,J+1\}} \mathbf{E} \left[\sum_{t'=t}^{T} \beta^{t'-t} u(s_{it'}, d_{it'}) + \epsilon_{d_{it'}} \mid \epsilon_{it}, s_{it}, d_{it} \right]$$

- State variables summarized in s_{it},

$$s_{it} = [X_t, E(H_t), V(H_t), \xi_t, r_t, MC_{it}]$$

Model: Value Function

Assuming that the transitions of the state variables are Markovian, write household's problem recursively,

$$V_t\left(s_{it}, \epsilon_{it}\right) = \max_{d_{it}}\{v_t(s_{it}, d_{it}) + \epsilon_{it}\} \tag{1}$$

where

$$\begin{aligned} v_t(s_{it}, d_{it}) \equiv & \; u(s_{it}, d_{it}) \\ + & \; \beta \int \sum_{s_{it+1}} V_{t+1}\left(s_{it+1}, \epsilon_{it+1}\right) q(s_{it+1} \mid s_{it}, d_{it}) dF(\epsilon_{it+1}) \end{aligned}$$

Model: CCP's and Finite Dependence

- Under assumptions common in discrete choice, differences in choice-specific value functions

$$
\begin{aligned}
v_t(s_{it}, d_{it} = j) - v_t(s_{it}, d_{it} = m) \\
&= u_j(s_{it}) - u_m(s_{it}) \\
&+ \beta \sum_{s_{it+1}} (u_k(s_{it+1}) - \log P_k(s_{it+1}))\, q(s_{it+1} \mid s_{it}, d_{it} = j) \\
&- \beta \sum_{s_{it+1}} (u_k(s_{it+1}) - \log P_k(s_{it+1}))\, q(s_{it+1} \mid s_{it}, d_{it} = m)
\end{aligned}
$$

which only depend on

- conditional choice probabilities, $E[P_k(s_{it+1}) \mid s_{it}]$, and
- transition probabilities, $q(s_{it+1} \mid s_{it}, d_{it})$

Back

Estimation: Stage 1

Estimate parameters of contamination equation

$$c_{jt} = H_j + \lambda \cdot IE_j + e_{jt}, \quad e_{jt} \sim N(0, \sigma_e), H_{j0} \sim N(0, \delta)$$

- H_j is unobservable $\longrightarrow$ use an EM algorithm (Dempster (1977))

E-step Given a guess of $\{\lambda, \sigma_e, \delta\}$, use data (c_{jt}, IE_j) to calculate posteriors on hazards using Bayesian updating formulas.

M-step Taking our estimate of hazard posterior as 'data', maximize likelihood of observing contaminants to get updated estimates of $\{\lambda, \sigma_e, \delta\}$.

EM Algorithm

E-Step: Given a guess of parameters at the k^{th} iteration, $\theta^{(k)} = \left[\lambda^{(k)}, \sigma_e^{(k)}, \delta^{(k)}\right]$,

- calculate $E(H_j)^{(k)}$ and $V(H_j)^{(k)}$ using Bayesian updating formulas

$$\begin{aligned} E(H_j)^{(k)} &= \left[\left(\delta^{(k)}\right)^{-1} + \frac{N_j}{\Sigma^{(k)}}\right]^{-1} \cdot \left(\left(\sigma_e^{(k)}\right)^{-1} \sum_{t=1}^{N_j}\left(c_{jt} - \lambda^{(k)} IE_j\right) + \left(\delta^{(k)}\right)^{-1} H_{j0}\right) \\ V(H_j)^{(k)} &= \left[\left(\delta^{(k)}\right)^{-1} + \frac{N_j}{\sigma_e^{(k)}}\right]^{-1} \end{aligned}$$

- This recovers a hazard level for each brownfield (of each neighborhood).
- Calculate likelihood of observing contamination given estimated posteriors

$$\begin{aligned} \int_{H_j} \log \ell(c_{jt} \mid H_j; \theta) dF(H_j) &= \int_{H_j} \log\left(\frac{1}{\sqrt{2\pi\sigma_e}}\, exp\left(\frac{(c_{jt} - H_j - \lambda \cdot IE_{jt})^2}{2\sigma_e}\right)\right) f(H_j) dH_j \\ &= -\frac{1}{2}\log(2\pi\sigma_e) - \frac{1}{2\sigma_e}\left[(c_{jt} - E(H_j) - \lambda \cdot IE_j)^2 + V(H_j)\right] \end{aligned}$$

EM Algorithm

E-Step: Given a guess of parameters at the k^{th} iteration, $\theta^{(k)} = \left[\lambda^{(k)}, \sigma_e^{(k)}, \delta^{(k)}\right]$,

- calculate $E(H_j)^{(k)}$ and $V(H_j)^{(k)}$ using Bayesian updating formulas

$$E(H_j)^{(k)} = \left[\left(\delta^{(k)}\right)^{-1} + \frac{N_j}{\Sigma^{(k)}}\right]^{-1} \cdot \left(\left(\sigma_e^{(k)}\right)^{-1} \sum_{t=1}^{N_j} \left(c_{jt} - \lambda^{(k)} IE_j\right) + \left(\delta^{(k)}\right)^{-1} H_{j0}\right)$$

$$V(H_j)^{(k)} = \left[\left(\delta^{(k)}\right)^{-1} + \frac{N_j}{\sigma_e^{(k)}}\right]^{-1}$$

- This recovers a hazard level for each brownfield (of each neighborhood).
- Calculate likelihood of observing contamination given estimated posteriors

$$\int_{H_j} \log \ell(c_{jt} \mid H_j; \theta) dF(H_j) = \int_{H_j} \log\left(\frac{1}{\sqrt{2\pi\sigma_e}} exp\left(\frac{(c_{jt} - H_j - \lambda \cdot IE_{jt})^2}{2\sigma_e}\right)\right) f(H_j) dH_j$$

$$= -\frac{1}{2}\log(2\pi\sigma_e) - \frac{1}{2\sigma_e}\left[(c_{jt} - E(H_j) - \lambda \cdot IE_j)^2 + V(H_j)\right]$$

M-Step: Maximize likelihood above to update $\theta^{(k)}$ to $\theta^{(k+1)}$.

Estimation: Stage 2

Estimate discrete choice problem

- Get state transition probabilities: $q(s_{it+1} \mid s_{it}, d_{it} = j)$
- Get conditional choice probabilities: $E[P_k(s_{it+1}) \mid s_{it}, d_{it} = j]$
- Recover utility parameters with MLE

$$L(\alpha) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} I_{[d_{it}=j]} \cdot \log \left(\frac{\exp\left(v_j(s_{it}) - v_m(s_{it})\right)}{1 + \sum_{\ell \neq m} \exp\left(v_\ell(s_{it})\right)} \right)$$

Back