

Housing II: Centralized Solutions

October 10, 2016

1 Discussion

- What is the trade off between targeting people who are at-risk for eviction, helping those who have already been evicted, or making long-term investments? One idea: If a household's wealth ever drops below a constant L , they suffer a bad outcome (like eviction) and are removed from the game. If a household's wealth ever reaches some constant H , they experience a good outcome (like moving to a lower-poverty neighborhood) and are removed from the game. Given that all households will eventually be removed from the game, how does a social planner with a budget disperse aid to maximize the good outcomes?
- How does a social planner minimize the long-run eviction rate given that those evicted are more likely to be evicted again?
- What should the social planner take into account? Both income AND recent changes. A family that drops from 20k to 10k probably is more at risk of eviction than one that is steady at 10k.
- The 30% rule is supposed to indicate that you're under financial stress with respect to housing if you're spending more than 30% of your income on housing. Does this rule account for expected shocks to income? Why did it arise?
- Let's ask the question: if we are given probabilities of individuals being evicted, and we have some budget, how do we allocate the aid? Economists often assume that the social planner has weights of how important individuals are.
- Is minimizing the eviction rate a reasonable goal? Well, maybe eviction is a proxy for the problems we actually care about. Do we care more about certain groups of people getting evicted (kids, poor families, ethnicities)?
- Is eviction a problem that people care about? By solving it, we could learn "these are the fires we put out first." Would that intuition be useful to someone who wants to work on this?
- We could mathematically compare different policies (help stabilize people who have been evicted vs. aid those at risk for eviction)

- When should we time housing related interventions relative to shocks/emergencies? We didn't specify what the interventions would be, how they would increase housing stability or welfare, or what is determined inside vs outside the model. Two issues to think about, using the example of housing vouchers: (1) an intervention like a housing voucher will affect the rents landlords charge to tenants, hence the social planner's budget constraint, and (2) agents might take undesirable actions to misrepresent their type (e.g. get evicted in order to get a housing voucher).
- Is there theory about ranking who to give funding to when you have multiple parameters (income, shock-factors, wealth)? Literature that insures people against shocks?
- What is the optimal design of housing assistance in the presence of temporary shocks? How long should housing assistance last – should it just be for a year to help people get back on their feet, or are states persistent enough that we want to make it permanent? There's quite a bit of econ literature on this topic (optimal design of insurance).
- “Hidden cost of being African American” – same income, but African American family not able to bounce back from shocks the same as white families. Suggests that income should not be the only factor.

2 Policies

We spend about 50 billion dollars on low-income housing, which is almost as much as what we spend on food stamps. This amount is not large compared to what we spend on housing overall (e.g., tax breaks for home owners).

There are 4 important federal low-income housing assistance programs: public housing; project-based vouchers; tenant-based vouchers; and the low-income housing tax credit (LIHTC).

These differ in how flexible they are, entities involved in operation (e.g., government, private landlords), and so on. We will note that in all of these, the rent formula is the same in that agents are not expected to spend more than 30% of their pre-tax income on rent. There are eligibility criteria that have to be met in order to qualify for these, and agents that end up using these programs are often well under the poverty line, even though the eligibility criteria can allow for a mix of income statuses.

1. Public Housing:

- \$8-10 billion per year, about 1.1 million households.
- Owned and operated by government (housing authority).
- Assistance is tied to a unit. You cannot take the “subsidy” with you if you move. Limited transfers are allowed within each city’s public housing system. Rules vary.
- Eligible if income below 80% area median income (AMI). Some housing authorities kick you out if your income rises above the threshold. Others do not.
- Tenants pay 30% pre-tax income towards rent. Note: what you pay is independent of what you get.
- Served on a first-come first serve basis, but there are times when priorities are given by location, veteran status or in case of an emergency (e.g., domestic abuse).

2. Project-based Vouchers:

- \$8-10 billion per year, about 1.2 million households.
- Like public housing, assistance is tied to a unit. Unlike public housing, the unit is privately owned and operated. Housing authority and owners collaborate on tenant selection.
- Eligible if $\leq 50\%$ AMI. Tenants lose voucher if income rises to where they no longer receive a subsidy.
- Tenants pay 30% pre-tax income towards rent.

3. Tenant-based Vouchers:

- Program was formerly known as Section 8, now “Housing Choice Vouchers.”

- \$16-18 billion per year, about 2.2 million households.
- Assistance tied to HOUSEHOLD, not unit. Voucher holders can change apartments and keep their subsidy.
- Used for rental on private market. Landlords and housing authority do not collaborate. Housing authorities allocate vouchers, landlords screen tenants.
- Eligible if $\leq 50\%$ AMI, same as project-based vouchers.
- Tenants pay 30% of pre-tax income towards rent. Total rent for unit capped at “Fair Market Rent”, a regional cap indexed to median rent and adjusted for apartment size. So the subsidy is limited, but up to the cap, tenant has no incentive to rent a cheaper apartment.
- Rented unit must meet certain quality standards. Few other restrictions (e.g. no restriction on neighborhood).
- In most (37) states, landlords can refuse to rent to a voucher holder. Discrimination against voucher holders prohibited in 13 states.

In addition to controlling allocation policy, housing authorities certify households every year (or sometimes every two years), checking income eligibility, criminal background, compliance with the lease, etc. More on this later.

4. Low-Income Housing Tax Credit (LIHTC):

- Relatively new program (since 1986). Largest driver of NEW “affordable” housing units.
- About \$7 billion per year.
- It’s a tax break for private developers who commit to renting new units at affordable rates. “Affordable” means that a household making 50-60% AMI would spend no more than 30% of its income on the unit. In other words, developer commits to charging no more than 15-18% AMI in rent. Each state is given a certain quantity of tax credits to award to new projects each year.
- Private developers own and operate these units, and also select tenants. Application process is totally decentralized. Housing authorities not involved, except when voucher holders rent LIHTC units.
- Importantly, rent is not indexed to tenant’s income. A very poor household pays a lot more for rent than in public housing or with a voucher.

2.1 Surrounding policy issues

- Program rationales:
 - Public housing began in 1937. Was touted as a macroeconomic stimulus plan and a subsidy to the construction industry.

- Began as a stepping stone for white working-class families, who would save money in public housing and then move to the suburbs (and buy a house with a federally subsidized mortgage).
 - Over time, became geared toward income redistribution and poverty alleviation. Also, shifted to minority tenants.
 - Some rationales are paternalistic: e.g. we know about benefits from decent housing that the households themselves don't take into account or realize at first.
- Quality of housing:
 - Mostly relevant for public housing. In the 1960s-1990s, public housing gained notoriety for poor housing conditions, crime, and mismanagement. Many forces contributed to this (anecdotally):
 - Some of it was legitimate administrative incompetence.
 - Housing authorities had discretion over where to build public housing, and for political and financial reasons most developments were built in the poorest neighborhoods.
 - The most notorious developments, such as Robert Taylor Homes and Cabrini Green in Chicago, were high-rises. This is now considered a poor model for affordable housing, because high-rises are expensive to maintain and for sociological reasons (community formation, monitoring, etc).
 - U.S. Department of Housing and Urban Development (HUD) provided inadequate funds to maintain existing buildings.
 - Income targeting: Should assistance go to the very poorest, or should we try to achieve mixed income developments and neighborhoods?
 - One the one hand, mixing incomes may create higher quality neighborhoods and avoid some of the historical problems associated with public housing.
 - One the other hand, every dollar that goes to a moderate-income household could have gone to a low-income household.
 - Currently HUD has targeting requirements for housing authorities:
 - Public Housing: 40% of units must go to households making $\leq 30\%$ AMI
 - Vouchers: 75% of voucher holders must make $\leq 30\%$ AMI
 Probably not a coincidence that the requirement for public housing is lower — everyone lives in the same place.
 - In the past, Congress has passed both weaker and stronger income targeting requirements. Depends on political climate.
 - Screening: Housing authorities have enormous discretion in screening applicants.
 - Some things must be checked by federal law: income, criminal background, immigration status.

- Many housing authorities also disqualify applicants with poor credit, rental, and/or employment histories. I don't know how much this actually happens. Cambridge claims not to do it. Many housing authorities claim the RIGHT to do it in their ACOPs (Admission and Continued Occupancy Policies).
- Tension here is analogous to income targeting — the applicants who might have the largest negative effects on other tenants might also have the highest need.
- In Cambridge at least, the real screening happens before applicants get to the top of the list (see below).

2.2 Allocation policies used by housing authorities

Focusing on allocation of public housing, since assistance is tied to a particular location and/or unit. Housing authorities use diverse rules for allocating public housing. Most systems involve the following features:

- Waiting lists: When you apply, you are placed on a set of waiting lists. These are usually first-come first-served within a priority group.
- Priority system: Certain types of applicants get housed before others. Most housing authorities have priorities for the following groups:
 - Applicants who live or work locally
 - Veterans
 - Emergency situations such as domestic violence or a no-fault eviction.
 - Lists are usually ordered lexicographically by priority and then by date of application. There are notable exceptions to this, such as NYC which has a rotating system for different priority groups.
- Development/location choice: Applicants usually have some choice over where they are assigned. Examples:
 - NYC: You first pick a borough. Later, you select from a menu of developments in that borough, and wait for that development.
 - Boston: You can select as many developments as you want (and have compatible apartments). You are placed on a waiting list for each development and are offered a unit from the first list where you reach the top.
 - Chicago: Public housing applicants can restrict housing offers to a certain part of the city. Only elderly and disabled applicants can pick specific developments.
 - Cambridge: You first select 3 developments. Later, you'll make a final selection and remain on the list for one development.

- Screening: applicants must verify income, criminal background, and immigration status. They must also provide any additional documentation — such as character references and landlord recommendations — required by the housing authority.
- Take-it-or-leave-it offer: applicants are penalized for rejecting offers. Usually, you have to start the application process over again. Most market design theorists think this leads to misallocation, since households may accept offers when they would prefer to wait for a better unit.
- List closures: many housing authorities close their waiting lists when they get too long.
 - Voucher waiting lists are usually closed. The Cambridge voucher list has been closed since 2008.
 - Chicago conducts lotteries to get on its waiting lists (not even for housing!).

One important fact learned from the Cambridge data is that the vast majority of applicants are removed from the waiting list without receiving housing assistance. Usually, this is because they don't respond to mail from the CHA. There is a limit to what I can learn from my data, but this disproportionately removes poorer households and other types of households that move frequently, which has a couple implications:

- Housing authorities may be screening out a group of extremely high-need applicants — households with the greatest housing instability.
- The size of a housing authority's waiting lists at any given time probably understates true demand for housing assistance (both because of list closures and because of removal).

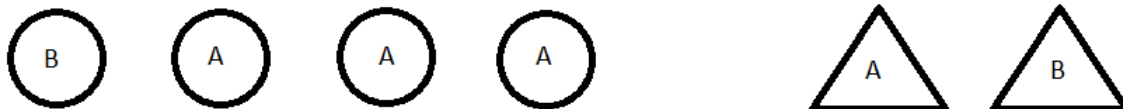
I would be surprised if these patterns were unique to Cambridge. Most housing authorities periodically contact applicants to confirm interest, although not all use mail as their primary method of contact.

Even apart from non-response, allocating housing at the local level harms households that move a lot. This is because housing authorities give locals priority, so households that move between cities effectively have to start over. This is less relevant in large cities like Boston, but potentially quite relevant in Cambridge, where lower-income households are rapidly being priced out of the city.

3 Notes on Leshno's "Dynamic Matching in Overloaded Waiting Lists"

3.1 Model

- Discrete Time: $t = 0, 1, 2, \dots$
- Each t , one item $x_t \in \{A, B\}$ becomes available. $Pr(x_t = A) = p_A$
- Some agents prefer A , others B . Denote their types α and β .
 - $u_\alpha(A) = 1; u_\alpha(B) = v, v < 1$. Symmetric for β .
 - Type α agents *match* objects of type A , and are *mismatched* to objects of type B .
 - p_α, p_β - frequencies of each type
- \mathcal{A}_t : set of agents on waiting list at time t . $\mathcal{A} \equiv \cup_{t \geq 0} \mathcal{A}_t$
- Agents enter the list by an exogenous process and remain on it until assigned an item
 - Key assumption: waiting list is overloaded. $|\mathcal{A}_t| \geq M$ for some $M \gg 0$
- All agents incur the same per-period waiting cost c



In the above example, we have agents on the left and two items arrive on the right hand side.

Definition: An *allocation* is $\mu : \{t > 0\} \rightarrow \mathcal{A}$ assigning each item x_t to agent a : $\mu(t) = a \in \mathcal{A}_t$.

- Must allocate item in the period it becomes available

3.1.1 Maximizing Welfare

Since total waiting cost is independent of who is assigned which item, the social planner simply wants to minimize the number of mismatched items. This simplifies welfare evaluation.

Definition: For a mechanism, let ξ_t denote the probability that the item in period t is misallocated. The *misallocation rate* ξ is defined

$$\xi = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \xi_t$$

Maximizing social welfare is equivalent to minimizing long-run misallocation, $(1 - v)\xi$.

Under the **full-information benchmark** – the social planner knows each applicant’s type – misallocation only happens if $x_t = A$ and there are no applicants of type A . This becomes less and less likely as the size of the waiting list grows, and in particular, if $p_\alpha = p_\beta = \frac{1}{2}$ and $\mathcal{A}_t \geq M \forall t$, then $\xi^{FB} \leq \frac{1}{2M}$.

With private information, the social planner must ask agents to report their types. WLOG, focus on direct revelation mechanisms.

Definition: A *deterministic dynamic matching mechanism* \mathcal{M} is defined by a state space \mathcal{S} , an initial state s_0 , and a transition function

$$\mathcal{T} : s_t \times \omega_t \mapsto s_{t+1} \times a \times k$$

where s_t is the state when period t begins, s_{t+1} is the state when it ends, and $\omega_t = (x_t, \omega_t^A) \in \{A, B\} \times \{\alpha, \beta\}^{\mathbb{N}}$ is the state of the world (the current item and the reported types of agents who haven’t been asked their types yet), $a \in \mathcal{A}$ is the agent assigned the item, and $k \in \mathbb{N}$ is the number of agents who reported their types during period t .

A simple example of a deterministic dynamic matching mechanism is pure first-come, first-served (FCFS): agents are ordered by arrival date, and each period the available item is assigned to the agent at the top of the list. The misallocation rate is $\xi^{Rand} = p_{AP\beta} + p_{BP\alpha}$. So, in the example above, when the first item arrives, we would assign it to the first agent and when the second item arrives, we would assign it to the second agent, resulting in the allocation below:

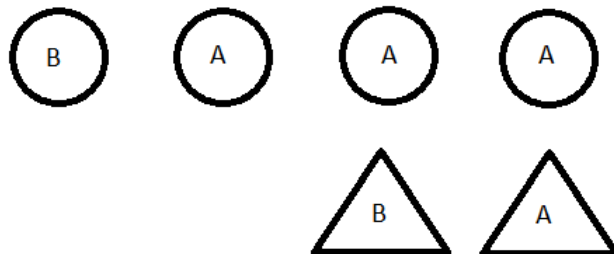


Figure 1: Simple Allocation Scheme

Note that the agents here are given a take it or leave it offer, and the second agent who is mismatched might refuse the item they are assigned, in which case the item would go to the third agent. This third agent might take this offer since they are further down in the queue and might be risk-averse, and fear that they won’t be assigned another item.

3.2 FCFS Buffer Queue

Consider the following direct revelation mechanism. This is not the optimal mechanism, but it improves upon random allocation and builds the apparatus used to derive the optimal mechanism. Agents are on a single waiting list and offered items in the order they applied. Importantly, they can decline offers and keep their place in line, and they know their place on the waiting list.

Consider the decision of a type α agent when offered an item. She will always accept a matching (type A) item, and the question is whether/when to accept a mismatched item. If she accepts a mismatched item, she gets $U_\alpha(B) = v$. The alternative is to decline the mismatched item and wait for the k -th matching item, where k is her place on the list. (Note that if the agent was offered a type B item and is not first in line, then those ahead of her are also type α , since they also declined that item) Thus, the expected payoff from waiting is

$$U_\alpha(\text{wait} \mid k) = 1 - c \frac{k}{p_A}$$

since her expected waiting time for a matched item is k/p_A . This yields a cutoff position, $K_\alpha = \lfloor p_A \frac{1-v}{c} \rfloor$, such that the applicant will decline a mismatched item if and only if $k \leq K_\alpha$. If the waiting list is long enough, the end of the list will never be reached. The condition is $M \geq 1 + \max\{K_\alpha, K_\beta\}$.

Given this decision rule, the state space of the mechanism can be represented with two *buffer queues* which contain the agents of each type who have declined mismatched offers. At any given time, only one buffer queue has any agents in it, yielding simple representation of the state space:

$$S = \{-K_\beta, \dots, -1, 0, 1, \dots, K_\alpha\}$$

The state follows an ergodic Markov Chain whose stationary distribution can be characterized in terms of transition probabilities that depend on the arrival rates of items and agents, and the resulting cutoff rules: $\{p_A, p_\alpha, K_\alpha, K_\beta\}$. In the simple case where items and agents are type-balanced, $p_A = p_\alpha = p$, the long-run misallocation rate is given by

$$\xi^{FCFS} = \frac{2p(1-p)}{(1-p)K_\alpha + pK_\beta + 1}$$

The numerator of this expression equals ξ^{Rand} . Allowing agents to reject mismatched offers reduces misallocation in proportion to K_α and K_β – the more agents that can be induced to decline mismatched offers, the lower the misallocation rate.

A couple comparative statics:

- As $c \rightarrow 0$, $\xi^{FCFS} \rightarrow 0$. If waiting costs are low, the state will almost never reach K_α or $-K_\beta$, which are the only states in which misallocation occurs.

- As $v \rightarrow -\infty$, $(1 - v)\xi^{FCFS} \rightarrow c$. As mismatched items become more undesirable, there are two countervailing forces. The probability of mismatch goes to zero, but the disutility from each mismatch grows. It happens that they cancel out.

3.3 Optimal Buffer Queue Policy

There is still misallocation in the FCFS buffer queue policy, and we'd like to reduce it further. Misallocation occurs due to temporary imbalances in supply (type A items) and demand (type α agents), so to reduce misallocation we should make expected waiting time as insensitive to these short-term fluctuations as possible. Some candidate approaches:

- Batches: wait until N items are available, and allocate them at once to N agents. This improves match quality but also increases waiting costs.
- Withhold waiting time information: if there are long-term imbalances in supply and demand ($p_A \neq p_\alpha$), some agents *should* take mismatched items.

Instead, Leshno's approach is to generalize the priority structure within buffer queues to incentivize rejecting mismatched offers.

A *buffer-queue policy* $\langle K, \varphi \rangle$ is given by a maximum number of agents in the buffer queue, K , and assignment probabilities $\varphi = \{\varphi(k, i)\}_{1 \leq i \leq k}$ such that $\sum_{i=1}^k \varphi(k, i) = 1$ for $k = 1, \dots, K$. So if an item arrives when there are k agents in the buffer queue, it is allocated to the agent in position i with probability $\varphi(k, i)$. For example, for first-come-first-served, $\varphi(k, i) = 1_{i=1}$, and for last-come-first-served, $\varphi(k, i) = 1_{i=k}$.

A *buffer-queue mechanism* $\mathcal{M} = \{K_\alpha, K_\beta, \varphi_\alpha, \varphi_\beta\}$ consists of two buffer-queue policies, $\langle K_\alpha, \varphi_\alpha \rangle$ and $\langle K_\beta, \varphi_\beta \rangle$.

A few results will be helpful for deriving the optimal mechanism:

- Welfare of an incentive-compatible mechanism \mathcal{M} depends *only* on K_α, K_β .
- The expected waiting time for an agent entering the α buffer queue at position k , w_k , only depends on ϕ_α (and not ϕ_β). So the two buffer queue policies can be analyzed separately.
- If agents have rational expectations about w_k , then incentive compatibility of \mathcal{M} is equivalent to $w_k \leq \bar{w}$ for $1 \leq k \leq K$ for each queue.

Leshno also makes the simplification that $p_A = p_\alpha = p$, which is the interesting case where there is no long-run supply and demand imbalance. With the above insights, the problem of finding the optimal mechanism can be formulated as follows: $\mathcal{M}^* = \langle K_\alpha^*, K_\beta^*, \varphi_\alpha^*, \varphi_\beta^* \rangle$, where each buffer queue policy solves

$$\max K \quad s.t. \quad w_k \leq \bar{w}$$

There are a few more steps in the argument for deriving the optimal mechanism, but the key insight is that K is maximized if w_k does not depend on k , i.e. $w_k = \bar{w} \forall k$. In other words, the expected waiting time for an agent who joins the buffer queue should not depend on the number of agents already in the queue. The optimal mechanism achieves this:

The **Load-Independent Expected Wait** ($LIEW_K$) buffer-queue policy is $\langle K, \varphi \rangle$ such that $w_k = \bar{w}$ for $k = 1, \dots, K$. In particular, $w_k = \frac{K+1}{2p}$.

Theorem: The optimal buffer-queue mechanism is $\mathcal{M}^* = \langle K_\alpha^*, K_\beta^*, \varphi_\alpha^*, \varphi_\beta^* \rangle$ with $K_\alpha^* = \lfloor 2p\bar{w} \rfloor - 1$, $K_\beta^* = \lfloor 2(1-p)\bar{w} \rfloor - 1$, $\varphi_\alpha^* = LIEW_{K_\alpha^*}$, and $\varphi_\beta^* = LIEW_{K_\beta^*}$. Long-run misallocation is given by

$$\xi^{OPT} = \frac{2p(1-p)}{(1-p)K_\alpha^* + pK_\beta^* + 1}$$

This is the same formula as for any buffer queue policy, just with maximal K^* 's.

3.4 Detail-free Mechanism – Service in Random Order (SIRO)

Implementing the LIEW buffer-queue policy is complicated – it requires adjusting $\varphi(k, i)$ to equate w_k for all k , and then randomizing. Furthermore, this policy can perform poorly if any of the following assumptions are violated:

- The model is misspecified
- The mechanism designer knows the parameters governing the model – v , c , and p – exactly.
- Applicants have correct beliefs about waiting times

Instead, it would be nice to have a nearly optimal mechanism with the following robustness properties:

- The buffer-queue policy is *scalable*, meaning it does not depend on specific parameters of the environment.
- The mechanism is *belief-free incentive compatible* (BF-IC), meaning it is incentive-compatible for an agent to report their true type regardless of their beliefs about the strategies of other agents. In other words, truth-telling must be a dominant strategy.
- Agents do not regret joining a buffer queue ex-post.

These properties can be obtained with a simple mechanism: Service in Random Order (SIRO). If the current size of the buffer queue is k , the item is offered with equal probability to all agents in the queue, regardless of their position within the queue: $\varphi(k, i) = \frac{1}{k}$

for $1 \leq i \leq k$. By giving equal weight to all positions, SIRO keeps the mechanism regret-free (agents at the top of the buffer queue never regret joining, since they would have been willing to join in position K) and minimizes the waiting time of the last agent in the queue subject to that constraint.

The robustness of SIRO offers practical benefits. Because SIRO draws agents from the buffer queue at random, it can be implemented easily without knowledge of model parameters. In simulations, it performs well relative to LIEW, though it does not achieve the optimal misallocation rate. It can also be implemented in more complex environments, e.g. with priority systems or where agents face heterogeneous waiting costs.