Aggregate Efficiency with Discrete Choice

David R. Baqaee
UCLA

Ariel Burstein
UCLA*

Abstract

This paper studies aggregate efficiency in general equilibrium models where households make a discrete choice about where they work. To measure the change in aggregate efficiency in response to technological change, we ask: "*what is the maximum reduction in total resources, given the new technologies, such that it is possible to make every household at least indifferent to their status-quo allocation?* If there is a single consumption good, then this measure is equivalent to asking: *once the winners compensate the losers, how much money is left over?* We characterize our measure of efficiency and show that, to a first-order approximation, its elasticity with respect to technology shocks is given by sales shares. This provides a justification for how multifactor productivity growth is measured in the national accounts. We also provide exact characterizations, showing how aggregate efficiency can be calculated by relying on cross-price elasticities of the location demand system, without explicitly making assumptions about the functional form of utility functions or the distribution from which tastes are drawn. We also contrast our measure with the common "expected utility" approach. We show that unless all households have the same preferences, then the two measures do not coincide even to a first-order.

*We thank Rodrigo Adao, Andrew Atkeson, Joao Guerreiro, Valentin Haddad, Oleg Itskhoki, Pablo Fajgelbaum, Gianluca Violante, and Pierre-Olivier Weill for their comments. We are especially grateful to Natalie Bau for her comments.

1 Introduction

In this paper, we study aggregate efficiency in general equilibrium models with discrete choice and heterogeneous preferences. Our approach differs from most of the rest of the literature in that we do not use a social welfare function or invoke ex-ante "expected utility" arguments to measure aggregate efficiency. Instead, we use the definition in Baqaee and Burstein (2025a). Suppose we start from some status-quo allocation, and technologies change. To measure the change in aggregate efficiency, we ask: "*what is the maximum reduction in total resources given the new technologies, such that it is possible to make every household at least indifferent to their status-quo allocation?*" If there is a single consumption good, this measure is related to the sum of compensating variations ("how much money is left over after winners compensate losers").

This question has several useful features. First, it has interpretable units expressed in terms of units of every good that are "left-over" once every household has been compensated for the change. Second, it depends only on the status-quo allocation, and primitives of the economy: technologies and preference relations. In particular, it does not require the economist to take a normative stance on Pareto-weights, inequality aversion, or cardinal properties of utility functions. This also means that our measure of efficiency has testable implications and can, at least in principle, be falsified. Both of these advantages are ultimately a consequence of the fact that our measure of aggregate efficiency is a counterfactual question about observables. Therefore, models that are observationally equivalent in terms of physical allocations, will assign the same number to the change in aggregate efficiency under our definition.

In the current version of the paper, we study only Pareto efficient models and competitive equilibria. That is, we abstract from distortions and externalities discussed by Fajgelbaum and Gaubert (2020) and Fajgelbaum and Gaubert (2025). Working with efficient models helps to clarify the differences between our approach and standard practice in the literature. However, extending the current analysis beyond Pareto efficient models is an important area for future work.

The structure of the paper is as follows. We set up the preferences and technologies in Section 2. Households have preferences over consumption and the region and/or industry in which they live and work. Production uses labor and intermediate inputs, and can accommodate input-output networks and trade but no agglomeration or congestion forces. In Section 3, we formally define our notion of aggregate efficiency. In this section, we show how to apply Theorem 1 from Baqaee and Burstein (2025a) in this setting, without taking a stance on how allocations are decentralized. In Section 4, we define a competitive equilibrium given preferences and technologies, and discuss how our measure of aggregate efficiency can be applied to analyze competitive equilibria specifically.

In Section 5 we prove a version of Hulten (1978) for economies with discrete choice. We show that, in competitive equilibria, the elasticity of aggregate efficiency to a productivity shock to a producer or region is equal to the sales of that producer or region divided by nominal GDP. This holds regardless of the underlying assumptions about utility functions and production functions. We show that our measure of aggregate efficiency coincides, to a first-order approximation, with measures of multifactor productivity growth (or the Solow residual) computed using a quality-adjusted labor input (as in, e.g., chapter 3 of the OECD's manual on measuring productivity). That is, our analysis provides a theoretical justification for those statistics in efficient economies with heterogeneous tastes and discrete choice.

Interestingly, we also show that this result does *not* typically apply to real GDP in such models, despite the fact that the economy is perfectly competitive. The intuition is that when people move from one job to another, their real wages adjust in a way that moves real GDP, but not aggregate efficiency. This is because any change in the real wage experienced by a marginal household that changes its choice is exactly offset by a change in the amenity value of that option for that household. Whereas the change in the real wage is measured in real GDP, the offsetting change in amenity value is not measured in real GDP. Since multifactor productivity growth statistics subtract the increase in real GDP due to compositional changes in the labor force, they correctly measure the increase in aggregate efficiency to a first-order.

In Section 6, we go beyond a first-order characterization by focusing on the special case of the model with a single traded consumption good. We provide exact (not just first-order) characterizations of the change in aggregate efficiency stated purely in terms of observable objects like sales shares, wage differentials, and migration decisions.

Crucially, we show that the market-level location demand system, which maps vectors of wages to population shares, can be used to calculate changes in aggregate efficiency without any explicit assumptions about the functional form of utility functions or the distribution of taste shifters in the population. In particular, this can be achieved by solving a differential equation in terms of cross-price elasticities of the location demand system, without having to specify or simulate the distribution of household tastes. We illustrate these results using the popular isoelastic location demand system discussed by Redding (2016) and Redding and Rossi-Hansberg (2017), which is typically derived assuming that household tastes follow a Frèchet distribution.

We end, in Section 7, by briefly contrasting our approach to the popular "expected utility" or "expected social surplus" function used in the literature. We highlight that our measure does not give the same answers except in the case where there is no household heterogeneity in tastes.

Relation to companion papers. Although this paper is self-contained, it has two companions. Baqaee and Burstein (2025a) provides a general framework for studying aggregate efficiency with heterogeneous agents. Many of the results in this paper are therefore applications of the general approach in that paper. In the other companion paper, Baqaee and Burstein (2025b), we apply the same framework to study efficiency costs of financial market incompleteness.

Relation to the literature. This paper is related to the large literature on discrete consumer choice with heterogeneous tastes, sometimes called random utility models. Many papers in this literature study partial equilibrium problems where agents chose a discrete good among many, taking prices and income as given, following McFadden (1981). The pioneers in this literature, like McFadden (1981) and Small and Rosen (1981), and Anderson et al. (1992), emphasized that consumer welfare in these models should be measured using compensating or equivalent variations of the individual agents. However, under the assumption that the indirect utility function is quasi-linear in income, it is known that there exists a closed-form formula, known as the "social surplus function," that can be used to calculate the sum of compensating variations. This social surplus function is the expected utility of the agents, under a particular cardinalizing assumption about taste shifters.

In the absence of quasi-linearity, calculating the sum of compensating variations typically requires resorting to simulation methods. See Hortaçsu and Joo (2023), and the references therein, for a recent textbook discussion. As discussed by Hortaçsu and Joo (2023), recent studies ignore compensating variations and directly use the social surplus function, the expected utility of agents under a cardinalizing assumption, as the starting point of their analysis.

A parallel and closely related literature is the spatial/occupational choice literature, which studies general equilibrium problems where agents choose a discrete location or job among many, taking prices but not their incomes as given. A counterpart to the social surplus function also exists in this literature, and this social surplus function is frequently used to conduct welfare analysis. However, this social surplus function, oftentimes called the "expected utility" function is not easily related to compensating variations of the individual agents.

In both of these literatures, the social surplus function is commonly used and justified on the grounds that it measures the "expected" utility to consumers. A recent example is Mongey and Waugh (2024), who consider an economy with an ex-ante stage where households have preferences over lotteries of their tastes. With this interpretation, one can study the welfare of this aggregate ex-ante agent before tastes are realized. They show that in such models, ex-post equilibrium allocations are inefficient since agents would want to insure each other against tastes before those tastes are realized. However, if such an ex-ante state does not exist, then the results of this type of analysis rest on an untestable cardinalizing assumption. That is, if households cannot make choices in the ex-ante stage that reveal their preferences about how they rank one set of taste parameters against another, then there is no conceivable choice data that can be used to recover the "ex-ante" expected utility function. Our approach is different since it does not rely on the existence of an ex-ante stage before tastes are known, nor does it rely on cardinal properties of utility functions.

Another related paper, which uses the social surplus or expected utility formulation, is Donald et al. (2023). They consider spatial models where households can be grouped into types, say by race or gender, and each type has heterogeneous tastes drawn from some type-specific distribution. As is standard practice in the literature, they compute a utilitarian social surplus function for each group. They then define aggregate welfare across groups using a social welfare function defined over the surplus function for each group. They decompose the first-order change in this notion of social welfare into different terms. The most important difference between our papers is that we ask a different question. As explained above, we do not use social welfare functions, or the utilitarian ("expected") social surplus function. Instead, we define aggregate efficiency in terms of resources savings after compensating transfers, and we characterize, both to a first-order and nonlinearly, in terms of familiar statistics of supply and demand curves. This means that the question we ask does not require taking a stance on interpersonal utility comparisons, it answers a counterfactual question about quantities, and is, at least in principle, falsifiable.

While the utilitarian approach is very common in the spatial literature, it is not universal. For example, Kim and Vogel (2020), who define aggregate welfare using the sum of workers' compensating variations, in a model in which workers choose among a discrete number of sectors with heterogeneous amenity values, non-labor participation and can be unemployed. They show that, a first order, the elasticity of aggregate welfare with respect to wages is given by income shares of each sector. Our measure of aggregate efficiency coincides with the sum of compensating variations in a simplified version of our model where there is a single consumption good. The first-order results in our paper complement the work in that paper, by proving an analogous result for aggregate efficiency in terms of productivity changes in general equilibrium with multiple goods and input-output networks. We also provide an exact characterization, in terms of observables, in the case when there is a single traded consumption good.

2 Environment

In this section we define the primitives of the economic environment: preferences, technologies, and resource constraints.

Households. Consider a collection of agents that choose among R different options, for concreteness, think of them as combinations of regions and industries. Agent h has preferences over its choice of options, $l_h \in R$, and a vector of consumption goods, $\boldsymbol{c}_h \in \mathbb{R}^{|R|}$, represented by the utility function $u_h(\boldsymbol{c}_h, l_h)$. We call l_h the location of household h, which includes both the region as well as the industry of occupation of h. Each worker has a unit endowment of time to work and can choose one, and only one, option:

$$\sum_r \mathbf{1}[l_h = r] = 1. \tag{1}$$

Denote the efficiency units of labor in location r by

$$L_r = \sum_h a_{hr} \mathbf{1}[l_h = r], \tag{2}$$

where a_{hr} is the efficiency units of household h if they choose to be located in r.

Technologies. Good r is produced according to the production technology

$$y_r = z_r F_r \left(\{x_{rj}\}_j, L_r \right), \tag{3}$$

where z_r is a Hicks-neutral productivity shifter for producer r, F_r is a constant returns production technology, and x_{rj} are intermediate inputs from $j \in R$.

Resource constraints. The resource constraint for good r is that consumption and intermediate input usage is less than production:

$$\sum_h c_{hr} + \sum_j x_{jr} \leq y_r. \tag{4}$$

Denote the consumption possibility set, given productivity shifters $\boldsymbol{z}$ and location of each household $\boldsymbol{l}$, by

$$\mathcal{C}(\boldsymbol{z}, \boldsymbol{l}) \equiv \left\{ \boldsymbol{c} \in \mathbb{R}^{N \times H} : \text{such that (2), (3), and (4) are satisfied for some } \{x_{ij}\} \right\}.$$

We sometimes suppress dependence of the consumption possibility set on $\boldsymbol{z}$ and $\boldsymbol{l}$.[1]

Below, we provide a toy example that we return to again and again to illustrate the general definitions and results.

Example 1 (Environment of the toy example). Consider an economy with two regions indexed by 1 and 2. Suppose there is a single consumption good, produced linearly from labor. Let z_1 and z_2 denote the productivity of labor in the two regions. Household h's utility function is[2]

$$u_h(c_h, l_h) = \epsilon_{h1} c_h \mathbf{1}[l_h = 1] + \epsilon_{h2} c_h \mathbf{1}[l_h = 2],$$

where ϵ_{hi} is the amenity value of choosing option i. Because there is a single consumption good, c_h is a scalar. The production technology in each region is

$$y_r = z_r L_r = z_r \sum_h a_{hr} \mathbf{1}[l_h = r],$$

where a_{hr} is the household-specific productivity of h in region r. The resource constraint is

$$\sum_h c_h \leq \sum_r y_r.$$

[1] Proportional shifts in the consumption possibility set can be achieved by scaling aggregate labor productivity. That is, consider the aggregate labor productivity shifter Z that affects the productivity of labor in every r simultaneously: $y_r = z_r F_r \left(\{x_{rj}\}_j, Z L_r \right)$. Since production functions have constant returns to scale, the feasible consumption sets scales proportionally with Z. That is, a doubling of Z doubles the consumption possibility set.

[2] We discuss more general preferences in Section 6.

Hence the production possibility set is defined by

$$\mathcal{C} = \left\{ \boldsymbol{c} \in \mathbb{R}^2 : \sum_h c_h \le \sum_r z_r \sum_h a_{hr} \mathbf{1}[l_h = r] \right\}.$$

In the rest of the paper, we refer to the economic environment in Example 1 as the *toy economy*. After stating each general result, we illustrate the intuition by applying it to the toy economy.

3 Aggregate Efficiency with Discrete Choice

Let the status-quo allocation be a collection of consumption vectors, $\boldsymbol{c}^0 = \{\boldsymbol{c}_h^0\}_h$, and location choices, $\boldsymbol{l}^0 = \{l_h^0\}_h$, for each household.

Definition 1 (Aggregate efficiency with location choice)**.** The change in aggregate efficiency relative to the status-quo is

$$A(\boldsymbol{c}^0, \boldsymbol{l}^0; \mathcal{C}) = \max \left\{ \phi : \text{there is } \boldsymbol{c} \in \phi^{-1}\mathcal{C}(\boldsymbol{z}, \boldsymbol{l}) \text{ and } (\boldsymbol{c}_h, l_h) \succeq_h (\boldsymbol{c}_h^0, l_h^0) \text{ for every } h \right\}.$$

We focus on log changes in A, starting from a point where the status-quo allocation is on the boundary of the status-quo feasible set: $\log A(\boldsymbol{c}^0, \boldsymbol{l}^0; \mathcal{C}(\boldsymbol{z}^0, \boldsymbol{l}^0)) = 0$. In this case,

$$\Delta \log A(\boldsymbol{c}^0, \boldsymbol{l}^0; \mathcal{C}) = \log A(\boldsymbol{c}^0, \boldsymbol{l}^0; \mathcal{C}) - \log A(\boldsymbol{c}^0, \boldsymbol{l}^0; \mathcal{C}(\boldsymbol{z}^0, \boldsymbol{l}^0)) = \log A(\boldsymbol{c}^0, \boldsymbol{l}^0; \mathcal{C}).$$

In words, $\Delta \log A$ is the maximum reduction in the consumption possibility set (e.g. achieved through a change in aggregate labor productivity) such that it is still possible to make every agent indifferent given their locations. If $\Delta \log A$ is a positive number, this effectively means we could divide every household's labor productivity by a factor A (shrinking the consumption possibility set by the same factor) and still make every household indifferent to the status-quo. Hence, our measure of aggregate efficiency, $\Delta \log A$, is a measure of the "left-over" resources once every agent has been compensated for the change at their given locations.

Example 2 (Aggregate Efficiency in toy economy). Let $\boldsymbol{c}^0$ and $\boldsymbol{l}^0$ denote the status-quo allocation in the toy economy. The change in aggregate efficiency is the maximum reduction in labor productivity for every household and every region by a common factor

$\Delta \log A$, so that the new possibility set is:

$$\left\{ \boldsymbol{c} : \sum_h c_h \leq \frac{1}{A} \sum_r z_r \sum_h a_{hr} \mathbf{1}[l_h = r] \right\},$$

such that it is still possible to keep every household indifferent to the initial allocation.

As discussed in Baqaee and Burstein (2025a), if there is a single consumption good, then $\Delta \log A$ coincides with the Kaldor-Hicks notion of efficiency. That is, $\Delta \log A$ can be interpreted as the ratio of aggregate nominal income to the sum of compensating variations for each household. The toy economy, which has a single consumption good, satisfies this property.

We now show that Theorem 1 from Baqaee and Burstein (2025a) can be extended to apply to this environment. To do so, we redefine the homothetized utility function of each agent.

Definition 2. Let $u_h : \mathbb{R}^{|R|} \times R \to \mathbb{R}$ denote a utility representation for agent h. The *homothteized utility function* $\tilde{u}_h : \mathbb{R}^{|R|} \times R \to \mathbb{R}$ is implicitly defined by

$$u_h(\frac{\boldsymbol{c}_h}{\tilde{u}_h}, l_h) = u_h(\boldsymbol{c}_h^0, l_h^0).$$

The homothetized utility function, $\tilde{u}_h$, is homogenous of degree one in consumption by construction. Given the homothetized utility functions, we define a Hicksian representative agent in this environment.

Definition 3. The *Hicksian representative agent* is an agent whose preferences are represented by

$$U(\boldsymbol{c}, \boldsymbol{l}) = \min_h \{\tilde{u}_h(\boldsymbol{c}_h, l_h)\},$$

where $\tilde{u}_h$ are homothetized utility functions.

Example 3 (Homothetized utility in the toy economy). The homothetized utility function of household h is

$$\tilde{u}_h(c_h, l_h) = \frac{\epsilon_{h1} c_h}{\epsilon_{hl_h^0} c_h^0} \mathbf{1}[l_h = 1] + \frac{\epsilon_{h2} c_h}{\epsilon_{hl_h^0} c_h^0} \mathbf{1}[l_h = 2],$$

Hence, $\tilde{u}_h(c_h, l_h)$ is simply equal to the growth in consumption if h stays in the same location, because $l_h = l_h^0$. However, if h moves locations, then $\tilde{u}_h$ is the growth in consumption times the ratio of the amenity in the new region relative to the status-quo region. Specifically, if ϵ_h in the status-quo region is larger than ϵ_h in the region h moved to, then $\tilde{u}_h$ is smaller than consumption growth, and the reverse is true if ϵ_h in the status-quo is smaller than ϵ_h in the region.

We are now in a position to extend Theorem 1 in Baqaee and Burstein (2025a), showing that calculating aggregate efficiency can be converted into an equivalent utility maximization problem.

Theorem 1 (Aggregate Efficiency by Utility Maximization with Discrete Choice)**.** *Define the value of $\mathcal{C}$ for the Hicksian representative agent to be*

$$V(\mathcal{C}) = \max_{\boldsymbol{c} \in \mathcal{C}(\boldsymbol{z}, \boldsymbol{l})} U(\boldsymbol{c}, \boldsymbol{l}).$$

The change in aggregate efficiency is equal to the value of $\mathcal{C}$ to the Hicksian representative agent:

$$\Delta \log A = \log V(\mathcal{C}).$$

Typically, assuming away corner solutions, the problem in Theorem 1 can equivalently be expressed as

$$V(\mathcal{C}) = \max_{\boldsymbol{c} \in \mathcal{C}} \left\{ \tilde{u}_h(\boldsymbol{c}_h, \boldsymbol{l}_h) : \tilde{u}_{h'}(\boldsymbol{c}_{h'}, l_{h'}) = \tilde{u}_h(\boldsymbol{c}_h, l_h) \text{ for every } h' \right\}.$$

In words, we maximizing $\tilde{u}_h$ for some specific household, subject to the additional constraints that $\tilde{u}_{h'}$ for every other household must be the same as $\tilde{u}_h$.

Below, we provide a simple example by returning to the toy economy.

Example 4 (Aggregate Efficiency in the toy economy). Let $\boldsymbol{c}^0$ and $\boldsymbol{l}^0$ denote the status-quo allocation in the two region example economy. Then, Theorem 1 implies that

$$\Delta \log A = \log \max_{\boldsymbol{c}} \min_{h} \left\{ \frac{\epsilon_{h1} c_h}{\epsilon_{hl_h^0} c_h^0} \mathbf{1}[l_h = 1] + \frac{\epsilon_{h2} c_h}{\epsilon_{hl_h^0} c_h^0} \mathbf{1}[l_h = 2] \right\},$$

subject to

$$\sum_h c_h = \sum_r z_r \sum_h a_{hr} \mathbf{1}[l_h = r].$$

Equivalently, fix some h, then

$$\Delta \log A = \log \max_{\boldsymbol{c}} \left\{ \frac{\epsilon_{hl_h} c_h}{\epsilon_{hl_h^0} c_h^0} \right\},$$

subject to

$$\sum_h c_h = \sum_r z_r \sum_h a_{hr} \mathbf{1}[l_h = r], \qquad \text{and} \qquad \frac{\epsilon_{hl_h} c_h}{\epsilon_{hl_h^0} c_h^0} = \frac{\epsilon_{hl_{h'}} c_{h'}}{\epsilon_{hl_{h'}^0} c_{h'}^0}, \qquad \text{(for every } h'\text{)}.$$

If we combine constraints, the solution is

$$A = \frac{\sum_r z_r \sum_h a_{hr} \mathbf{1}[l_h = r]}{\sum_h \frac{\epsilon_{hl_h^0}}{\epsilon_{hl_h}} c_h^0}. \tag{5}$$

This expression is directly interpretable. The numerator is total production of the consumption good. The denominator is the sum of $\frac{\epsilon_{hl_h^0}}{\epsilon_{hl_h}} c_h^0$ over all h. The number $\frac{\epsilon_{hl_h^0}}{\epsilon_{hl_h}} c_h^0$ is the minimum consumption agent h needs to be given in location l_h to ensure indifference relative to the status-quo.

Notice that if there is no heterogeneity in household tastes for locations, i.e. $\epsilon_{hr} = \epsilon_{hr'}$ for each h, then (5) is simply the ratio of the total quantity of the consumption good given the new technologies relative to the status-quo.

4 Competitive Equilibrium with Discrete Choice

We now define the competitive equilibrium. Each household h maximizes utility $u_h(\boldsymbol{c}_h, l_h)$ subject to two constraints. The first constraint, (1), implies that household can only choose one of the discrete options. The second constraint is the budget constraint

$$\sum_r p_r c_{hr} = \sum_r w_r a_{hr} \mathbf{1}[l_h = r],$$

which states that consumption expenditures must be financed by labor income in the location the household chooses — where the nominal wage per efficiency unit of labor in region r is w_r.

Firm r chooses inputs x_{rj} and labor inputs L_r to maximize profits

$$p_r y_r - \sum_j p_j x_{rj} - w_r L_r$$

subject to its production technology, (3), taking prices as given.

We now define general equilibrium with discrete choice.

Definition 4 (Equilibrium with Discrete Choice)**.** An equilibrium is a collection of consumption vectors, output vectors, intermediate input choices, location choices, prices, and wages such that each household chooses consumption and location to maximize utility subject to their budget constraint and (1); firms choose output and inputs to maximize profits taking prices as given; and all resource constraints, (2) and (4), are satisfied.

We illustrate competitive equilibrium inside the toy economy.

Example 5 (Competitive Equilibrium in the toy economy). Returning to the toy economy, the budget constraint of each household is

$$c_h = w_1 a_{h1} \mathbf{1}[l_h = 1] + w_2 a_{h2} \mathbf{1}[l_h = 2],$$

where we have set the price of the single consumption good to be the numeraire. In the competitive equilibrium, the real wage per efficiency unit in each region is equal to labor productivity in that region: $w_r = z_r$.

The only variable left to be determined in equilibrium is each household's choice of location. For elegance, assume that there is a unit mass of households continuously distributed according to the density function $g(\epsilon_1, \epsilon_2, a_1, a_2)$, where ϵ_1 and ϵ_2 are amenity values and a_1 and a_2 are location-specific productivities (all strictly positive). To pin down the equilibrium allocation, rank households by the ratio of $x_h = (a_{h2}/a_{h1}) \times (\epsilon_{h2}/\epsilon_{h1})$. Define the cut-off value

$$x^* = \frac{z_1}{z_2}.$$

Household h locates in region 1 if, and only if, $x_h < x^*$, otherwise, household h locates in region 2. In equilibrium, the efficiency units of labor in region 1 is

$$L_1 = \int\int\int\int a_1 \mathbf{1}\left[\frac{a_2}{a_1}\frac{\epsilon_2}{\epsilon_1} \leq x^*\right] g(\epsilon_1, \epsilon_2, a_1, a_2) d\epsilon_1 d\epsilon_2 da_1 da_2,$$

where g is the joint density of taste and productivity shifters. A similar expression holds for L_2. In the competitive equilibrium, household h in region l_h consumes $z_h a_{hl_h}$ units of the consumption good.

Consider the special case where there is no heterogeneity in productivity, say $a_{h1} = a_{h2} = 1$ for every h. In this case, denote the density of $x_h = \epsilon_{h2}/\epsilon_{h1}$ by f. In this simple case, the integral above simplifies to just

$$L_1 = \int_0^{x^*} f(x) dx,$$

which is the fraction of households for whom $x_h = \epsilon_{h2}/\epsilon_{h1}$ is less than $x^* = z_2/z_1$.

Having defined competitive equilibrium with discrete choice, we now consider how aggregate efficiency changes in response to productivity shocks assuming competitive equilibrium. Let $\boldsymbol{l}(\boldsymbol{z})$ be equilibrium location decisions by each household given productivity shifters $\boldsymbol{z}$. Define the consumption possibility set by $\mathcal{C}(\boldsymbol{z}, \boldsymbol{l}(\boldsymbol{z}))$. The change in aggregate efficiency, given some status-quo allocation, can then be obtained by Theorem 1 applied to this consumption possibility set.

In the next two sections, we provide first-order and nonlinear characterizations of the change in aggregate efficiency comparing, $\mathcal{C}(\boldsymbol{z}, \boldsymbol{l}(\boldsymbol{z}))$, to the status-quo competitive equilibrium allocation, $(\boldsymbol{c}(\boldsymbol{z}^0), \boldsymbol{l}(\boldsymbol{z}^0))$.

5 First Order Results

In this section, we consider first-order changes in aggregate efficiency in response to first-order changes in productivity shifters $\boldsymbol{z}$ assuming the economy is in a competitive equilibrium. We show that changes in aggregate efficiency are given by sales shares in the status-quo, and one does not need to solve for how allocations respond to the shocks because of the envelope theorem.

To do so, index productivity shifters by a scalar t.[3] We interpret the status-quo allocation to be the allocation generated by a competitive equilibrium with $\boldsymbol{z}(0)$. We then consider how aggregate efficiency, $A(t)$, changes as t changes, always relative to the status-quo allocation associated with $\boldsymbol{z}(0)$.[4] Throughout the paper, for any variable X, we use the shorthand dX to denote infinitesimal changes $dX \equiv \frac{dX}{dt}dt$. For non-infinitesimal changes, we write $\Delta X = X(t) - X(0) = \int_0^t dX$. When there is no ambiguity, we suppress the dependence of ΔX and dX on t.

To state the result, denote sales of r relative to aggregate nominal consumption by

$$\lambda_r(t) = \frac{p_r(t)y_r(t)}{\sum_r \sum_h p_r(t)c_{hr}(t)}.$$

This statistic is popularly called the Domar weight of r. Note that, in general, $\sum_r \lambda_r > 1$ if there are intermediate inputs. This is because if there are intermediate inputs, then the sum of sales is greater than the sum of consumption expenditures.

[3] The scalar t simply indexes primitives, and may or may not correspond to time.

[4] With some abuse of notation, this means that $A(t) = A(\boldsymbol{c}(\boldsymbol{z}(0)), \boldsymbol{l}(\boldsymbol{z}(0)); \mathcal{C}(\boldsymbol{c}(\boldsymbol{z}(t)), \boldsymbol{l}(\boldsymbol{z}(t))))$.

Proposition 1 (First-Order Efficiency Change with Discrete Choice)**.** *To a first-order approximation in productivity shocks, the change in aggregate efficiency is*

$$\Delta \log A \approx \sum_r \lambda_r(0) \Delta \log z_r,$$

where the approximation error is $(\Delta \log z_r)^2$*.*

Intuitively, the change in aggregate efficiency is "as-if" households do not change locations in response to the productivity shock. The reason is that any change in consumption caused by a change in location is exactly offset by a change in amenity value for the marginal household that changed locations. Hence, such effects can be ignored to a first-order. We illustrate Proposition 1 by applying it to the toy economy.

Example 6 (Change in aggregate efficiency in toy economy). In the two region economy, changes in aggregate efficiency, to a first order, satisfy

$$\Delta \log A \approx \frac{w_1 L_1}{w_1 L_1 + w_2 L_2} \Delta \log z_1 + \frac{w_1 L_1}{w_1 L_1 + w_2 L_2} \Delta \log z_2,$$

where $w_i L_i$ are evaluated at the status-quo allocation. Hence, an increase in productivity in region r raises aggregate efficiency by that region's Domar weight (or equivalently, its share of aggregate income).

We now contrast changes in aggregate efficiency with changes in real GDP. It is well known that in a competitive equilibrium without discrete choice, if labor (and other primary factors like land) are inelastically supplied, then the elasticity of real GDP to productivity shocks is given by Domar weights (see Hulten, 1978). Proposition 1 shows that this result applies to aggregate efficiency in our environment, but we show that it need not apply to real GDP. First, define real GDP.

Definition 5. We define the (infinitesimal) change in real GDP in the usual way, as a share-weighted sum of changes in final consumption quantities:

$$d \log Y(t) = \sum_r \sum_h \frac{p_r(t) c_{hr}(t)}{\sum_r \sum_h p_r(t) c_{hr}(t)} d \log c_{hr}.$$

Non-infinitesimal changes are defined by integrating $d \log Y$.

We illustrate the difference between $\Delta \log A$ and $\Delta \log Y$ using the toy economy.

Example 7 (Real GDP versus aggregate efficiency in the toy economy). In the two region economy, changes in real GDP are simply the change in total consumption:

$$d\log Y = \sum_r \sum_h \frac{p_r c_{hr}}{\sum_r \sum_h p_r c_{hr}} d\log c_{hr} = \frac{d\sum_r \sum_h c_{hr}}{\sum_r \sum_h c_{hr}} = d\log \sum_r \sum_h c_{hr},$$

since there is only one consumption good with the same price in both regions $p_1 = p_2$. That is, the change real GDP (in logs) is simply the log change in the aggregate quantity of the consumption good.[5]

The response in real GDP, even with more than two regions, to a first order, is given by

$$\Delta\log Y \approx \Delta\log A + \frac{1}{\sum_{r'} w_{r'} L_{r'}} \int \sum_{i,j\in R} \left[w_i a_i - w_j a_j\right] \Delta L_{j\to i}(\boldsymbol{a}) f(\boldsymbol{a}) d a, \tag{6}$$

where $\Delta L_{j\to i}(\boldsymbol{a})$ is the share of agents with productivity vector $\boldsymbol{a}$ who move from region j to region i and $f(\boldsymbol{a})$ is the marginal distribution of productivity vectors $\boldsymbol{a}$. In words, real GDP is the change in aggregate efficiency plus the change in real income for those households that move between regions (relative to nominal GDP). From this equation, it is clear that $\Delta\log Y$ coincides with the change in aggregate efficiency, $\Delta\log A$, if there is no mobility, $\Delta L_{j\to i} = 0$.

The easiest way to see the difference is to imagine there is a single household in the economy living in location 1 but indifferent between living in location 1 or 2. For example, the real wage in location 1 is higher than in location 2, but the amenity value of living in 2 is higher than 1. In this case, an infinitesimal increase in the productivity in location 2 causes the household to move to location 2. Since there is an infinitesimal increase in productivity, efficiency rises by an infinitesimal amount. The real wage falls discretely when the agent migrates, but the amenity value rises discretely to offset that. However, real GDP falls discretely by the gap between the real wage in location 2 and location 1 once the household migrates.

In the two-region example, we can rewrite this as

$$\Delta\log Y \approx \Delta\log A + \int\int \frac{\left[w_1 a_1 - w_2 a_2\right] g\left(\frac{z_2}{z_1}\frac{a_1}{a_2} | a_1, a_2\right) d\left[\frac{z_2}{z_1}\frac{a_1}{a_2}\right]}{w_1 L_1 + w_2 L_2} f(a_1, a_2) d a_1 d_2,$$

[5] For this toy economy, notice that this first-order approximation is exact:

$$\Delta\log Y(t) = \int_0^t d\log Y = \int_0^t d\log \sum_r \sum_h c_{hr} = \Delta\log \sum_r \sum_h c_{hr}.$$

where $g(\cdot|a_1, a_2)$ is the conditional density of ϵ_1/ϵ_2 given a_1 and a_2, and f is the density of the household-by-location specific productivity vector (a_1, a_2). To make this formula easier to see through, consider two extreme cases. In the first case, suppose there is no household-by-location productivity heterogeneity. In the second extreme case, suppose that there is no household-by-location taste heterogeneity.

Consider the first extreme. When there is no household-by-location productivity heterogeneity, $f(a_1, a_2)$ is a Dirac delta function, without loss of generality, say at $(a_1, a_2) = (1, 1)$. In this case,

$$\Delta \log Y \approx \Delta \log A + \frac{[w_1 - w_2]\, g(\frac{z_2}{z_1}) d \left[\frac{z_2}{z_1}\right]}{w_1 L_1 + w_2 L_2}.$$

The term $g(\frac{z_2}{z_1}) d \left[\frac{z_2}{z_1}\right]$ is the mass of households that move between 1 and 2 in response to the shock. The change in real GDP coincides with the change in aggregate efficiency only if either there are no households at the cut-off, $g(z_2/z_1) = 0$, or if there is no difference between real wages in the two regions $z_1 = z_2$.

Now consider the second extreme. When there is no household-by-location taste heterogeneity, say $\epsilon_1 = \epsilon_2$, the marginal households are ones for whom $w_1 a_1 = w_2 a_2$. Hence, the general expression simplifies to just

$$\Delta \log Y \approx \Delta \log A$$

because when the marginal household relocates from one location to another, their real income does not change.

We end this section by illustrating the relationship between $\Delta \log A$ and multifactor productivity using the toy economy.

Example 8 (Multifactor productivity versus aggregate efficiency in the toy economy). Define multifactor productivity, following Solow (1957) and OECD (2001), as

$$d \log A^{MFP} = d \log Y - \sum_r \frac{w_r}{\sum_{r'} w_{r'} L_{r'}} dL_r.$$

The second term is the change in the "quality-adjusted" labor input. The intuition here is to separate changes in output due to technology shocks from changes in output due to changes in the skill composition of workers. Substituting (6) into this expression and rearranging yields

$$\Delta \log A^{MPF} \approx \Delta \log A.$$

In words, the change in aggregate efficiency, as measured by $\Delta \log A$, coincides with the change in multifactor productivity growth, as measured by $\Delta \log A^{MFP}$, to a first-order approximation. However, as we show in next section, this equivalence between $\Delta \log A$ and the Solow residual breaks down beyond the first-order.

---

6 Nonlinear Characterization of Efficiency in Terms of Observables

In this section, we go beyond the first-order approximation in Proposition 1 and characterize nonlinear changes in aggregate efficiency. We do so for a special case but we expect many of the steps generalize beyond the simple case that we consider here. The key idea in this section is that changes in aggregate efficiency can be expressed without reference to unobservable parts of the model like the distribution of taste shifters. Instead, we characterize aggregate efficiency purely in terms of familiar statistics of demand curves and other observables.

This means that conditional on the same observables (like movements from region i to region j given a shock), our aggregate efficiency measure is the same, and the economist does not have to estimate or take a stance on issues like the underlying distribution of taste shocks or the functional form of the utility function (e.g. are the taste shocks additive or multiplicative).

We begin this section by describing a special case of the general environment. We then state our characterization result.

6.1 Specializing the Environment

The model we focus on is a generalization of the toy economy, but allowing for more than two options and more general utility functions.

Preferences. There is a unit mass of households with preferences $u_h(c, l)$ over a single consumption good and locations (recall that a "location" could be in space or in occupations). Following Theorem 1, we use "homothetized" utility functions to calculate aggregate efficiency. If household h locates in l_h, the homothetized' utility function of h has the form

$$\tilde{u}_h(c_h, l_h) = \epsilon_{hl_h} c_h,$$

where ϵ_{hl_h} is a household-by-location shifter capturing tastes for living in location l_h. To see this, note that without loss, we can write $u_h(c_h, l_h) = \sum_r f_{hr}(c_h)\mathbf{1}[l_h = r]$, where f_{hr} is a household-by-region specific strictly increasing function. Applying the definition of the homothetized utility function and inverting gives: $\tilde{u}_h(c_h, l_h) = \sum_r \left[f_{hr}^{-1}(u_{h0})\right]^{-1} c_h\mathbf{1}[l_h = r] = \sum_r \epsilon_{hr}c_h\mathbf{1}[l_h = r] = \epsilon_{hl_h}c_h$, where $\epsilon_{hr} \equiv \left[f_{hr}^{-1}(u_{h0})\right]^{-1}$. This also pins down the taste parameter for the status-quo location $\epsilon_{hl_h^0} = 1/(c_h^0)$. As long as we do not put any assumptions on the distribution of ϵ_{hr}, this is fully general. The only restriction we place on the distribution of household h's taste parameters, $\boldsymbol{\epsilon}_h$, is that they are distributed according to some continuous multivariate distribution.

Households choose the location they live l_h to maximize utility $u_h(c_h, l_h)$ subject to the budget constraint

$$c_h = \sum_{r \in R} w_r\mathbf{1}[l_h = r],$$

where w_r is the real wage associated with option r. The real wage is unambiguous to define since there is only one consumption good.

Production and resource constraints. The single consumption good is produced linearly from labor. Production in region r is

$$y_r = z_r \sum_h \mathbf{1}[l_h = r].$$

We abstract from household-region-specific productivity differences for simplicity. The equilibrium real wage in region r is $w_r = z_r$. Goods market clearing requires that total consumption by all households equals total production by all producers

$$\sum_h c_h = \sum_r y_r.$$

6.2 Characterization

Index productivities $\boldsymbol{z}(t)$ by a scalar t and assume that the status-quo corresponds to $\boldsymbol{z}(0)$. We are interested in understanding how aggregate efficiency evolves as a function of t. For concreteness, we refer to t as "time", though this is simply a label for the scalar indexing primitives, and the economic environment is static.

Proposition 2 (Exact Characterization of Efficiency). *Theorem 1 implies that aggregate effi-*

ciency, $A(t)$, is equal to

$$A(t) = \frac{\sum_r z_r(t) \sum_h \mathbf{1}[l_h(t) = r]}{\sum_h \frac{\epsilon_{hl_h^0}}{\epsilon_{hl_h}} c_h^0}, \tag{7}$$

where

$$l_h(t) = \arg\max_l \{u_h(w_l(t), l)\} = \arg\max_l \{u_h(z_l(t), l)\}. \tag{8}$$

Proposition 2 provides a relatively straightforward way to calculate aggregate efficiency conditional on knowing the status-quo allocation and the primitive preferences and technologies. First, compute location decisions in the competitive equilibrium, given $\boldsymbol{z}(t)$, using (8) and then evaluate (7) for $A(t)$. The proof and intuition for this result are identical to the derivation of (5) in Example 4.

As we discuss at the end of this section, Proposition 2 can be extended to allow for different household types, say with different productivities in different locations, relatively easily. Proposition 2 can also be applied to characterize efficiency for subsets of households. Suppose we apply (7) and (8) to only a subset of households $H' \subset H$. In this case, $A(t)$ measures the maximum contraction in the consumption possibility set, among the households in H', holding fixed all other households' consumption choices, such that it is possible to keep every household in H' indifferent. This is useful for defining and studying efficiency by group: for example skill level, or initial location, etc.

Although Proposition 2 is useful for computing $A(t)$ given a fully specified structural model, it suggests that we need to model both the specific form of utility functions $u_h(c_h, l)$ as well as relative $\epsilon_{hr}/\epsilon_{hl_h^0}$ at the individual level in order to calculate aggregate efficiency. We now provide an alternative characterization, which shows that we do not need direct knowledge of either the utility functions or the distribution of ϵ's and can get by if we know the induced aggregate demand for each location as a function of the vector of real wages.

To do this, denote the share of households living in region r at time t by

$$L_r(t) = \int \mathbf{1}[l_h(t) = r] dh,$$

where $l_h(t)$ is the location of household h in the competitive equilibrium at time t. That is, $l_h(t)$ solves the problem in (8). To characterize nonlinear changes, define the set of households that move from r to r' to be

$$L_{r \to r'}(t) = \int \mathbf{1}\left[l_h(t_0) = r\right] \mathbf{1}\left[l_h(t) = r'\right] dh.$$

The sales share of location r, relative to GDP, is given by

$$\lambda_r(t) = \frac{w_r(t)L_r(t)}{\sum_{r'} w_{r'}(t)L_{r'}(t)}.$$

The next proposition provides a second-order approximation of aggregate efficiency. We express the change in aggregate efficiency in terms of primitive shocks and observables, like sales shares, wage differentials, and the number of households that switch locations without direct reference to the distribution of taste parameters.

Proposition 3 (Second-Order Approximation of Aggregate Efficiency)**.** *At $t = 0$, the change in aggregate efficiency is, to a second-order approximation, given by*

$$\Delta \log A \approx \sum_r \left[\lambda_r + \frac{\Delta \lambda_r}{2}\right] \Delta \log z_r + \frac{1}{2} \sum_{r,r'} \frac{[w_{r'} - w_r]}{\sum_{r'} w_{r'} L_{r'}} \Delta L_{r \to r'} \left[\Delta \log z_r - \sum_i \lambda_i \Delta \log z_i\right]$$

The first summand, $\sum_r \lambda_r \Delta \log z_r$, is just the first-order approximation. The second summand, $1/2 \sum_r \Delta \lambda_r \Delta \log z$ are the nonlinearities terms emphasized by Baqaee and Farhi (2019) — if Domar weights λ_r rise for options where productivity $\Delta \log z_r$ grew, then this nonlinearity amplifies the beneficial impacts of positive productivity shocks (and mitigates the harmful effects of negative productivity shocks). In other words, positive productivity shocks in region r are more beneficial if the sales share of r grows with the shock. The sum of $\sum_r \lambda_r \Delta \log z_r$ and $1/2 \sum_r \Delta \lambda_r \Delta \log z$ equals multifactor productivity growth to a second-order.

The final summand is new relative to Baqaee and Farhi (2019) and cause $\Delta \log A$ to not equal multifactor productivity growth to a second-order. This term states that migrations from r to r' that raise real wages for households that move reduce aggregate efficiency if productivity growth in r is less than average. Intuitively, this is because households leaving r are relatively less willing to live in other locations as the shock grows, and hence the gains they experience from relocating are smaller.

A crucial fact about Proposition 3 is that it shows what information is required to be able to calculate changes in aggregate efficiency, to a second-order, in the data. If the economist can observe initial sales and wages, and can, through instruments, identify the change in sales and migration decisions induced by shocks, then calculating $\Delta \log A$, up to a second order, is simple calculation that does not require any additional assumptions about preferences.

Proposition 3 is useful it relates aggregate efficiency to statistics of the equilibrium allocation. However, applying it requires knowledge of migration decisions. To fill in this missing step, define the vector-valued function $\boldsymbol{L}(\boldsymbol{w})$ which maps vectors of real wages $\boldsymbol{w}$ to the share of households that choose each option. In particular, $\boldsymbol{L}(\boldsymbol{w})$ aggregates equilibrium individual location choices given by (8). We refer to $\boldsymbol{L}$ as the location demand system.

To make $\boldsymbol{L}(\boldsymbol{w})$ more concrete, we provide one example of a location demand system below. We do not impose this functional form for our theoretical results, but simply use it for illustration.

Example 9 (Isoelastic Demand System). Suppose that the location demand system is

$$L_r(\boldsymbol{w}) = \frac{B_r w_r^{\theta}}{\sum_{r'} B_{r'} w_{r'}^{\theta}}, \tag{9}$$

where $B_r > 0$ are some scalar demand shifters. This demand system is very popular in the spatial literature (see e.g. Redding 2016, and the literature review in Redding and Rossi-Hansberg 2017). In Section 7 we return to this demand system and discuss the distribution of household-level taste parameters that can rationalize it. Since our results do not depend directly on the distribution of taste shifters, beyond their effect on (9), we do not state the underlying taste parameters in this section.

The following proposition uses the cross-price elasticities of the location demand system to pin down equilibrium allocations. This can then be fed into Proposition 3 to conduct counterfactuals.

Proposition 4 (Characterization of Equilibrium Allocation)**.** *In equilibrium, the number of movers between options satisfies*

$$\frac{dL_{r\to r'}}{dt} = \frac{\partial L_r}{\partial \log w_{r'}} \left[\frac{d\log w_r}{dt} - \frac{d\log w_{r'}}{dt} \right] \mathbf{1}\left[d\log w_{r'} \geq d\log w_r \right].$$

The real wage in each region is equal to the productivity of that region $w_r(t) = z_r(t)$. The number of people choosing each option r is given by the location demand system $L_r(t) = L_r(\boldsymbol{w}(t))$, and the Domar weight of each region satisfies

$$\lambda_r(t) = \frac{w_r(t) L_r(t)}{\sum_{r'} w_{r'}(t) L_{r'}(t)}.$$

Proposition 4 does not impose a functional form on the location demand system beyond differentiability. Proposition 4 pins down equilibrium allocations at t in terms of the primitive productivity shocks $\boldsymbol{z}(t)$ and the location demand system. Therefore, given knowledge of the location demand system, one can easily solve for aggregate efficiency, without making direct assumptions about the distribution of household tastes. Moreover, whereas real wages, sales, and population shares in each location at t are ultimately only a function of t, the number of switchers between r and r' depends both on $\boldsymbol{z}(t)$ and $\boldsymbol{z}(0)$.

The following example illustrates using the isoelastic demand system.

Example 10 (Switchers in Isoelastic Location Demand System). Assume that the isoelastic location demand system in (9). The share of switches from r to r' satisfies the differential equation

$$\frac{dL_{r\to r'}}{dt} = \theta L_r(t)L_{r'}(t)\left[\frac{d\log w_{r'}}{dt} - \frac{d\log w_r}{dt}\right]\mathbf{1}\left[\frac{d\log w_{r'}}{dt} \geq \frac{d\log w_r}{dt}\right].$$

The share of households switching from r to r' is positive only if the real wage in r' is rising faster than in r. The intensity of the flow is increasing in the elasticity of location demand, and the initial share of the population in r and r'. Plugging this equation into Proposition 4 gives a characterization of changes in aggregate efficiency that depend only on cross-price elasticities of the location demand system, initial expenditures, and primitive shocks.

Extension with household-by-location productivity differences In this section, we assumed for simplicity that there is no household-by-location productivity differences. Our results can be extended to cover such an environment. For example, if there are idiosyncratic productivity differences, then Proposition 2 must be modified in the following way.

Proposition 5 (Exact Characterization of Efficiency with Idiosyncratic Productivity)**.** *Theorem 1 implies that aggregate efficiency, $A(t)$, is*

$$A(t) = \frac{\sum_r z_r(t)\sum_h a_{hr}\mathbf{1}[l_h(t) = r]}{\sum_h \frac{\epsilon_{hl_h^0}}{\epsilon_{hl_h}} c_h^0},$$

where a_{hr} is household h's productivity in location r. The location choices that solve this problem are given by

$$l_h(t) = \arg\max_l\{u_h(w_l(t)a_{hl}, l)\} = \arg\max_l\{u_h(z_l(t)a_{hl}, l)\}.$$

The rest of the results can also be extended along similar lines, but we do not state them here for brevity. Intuitively, we would apply each result conditioning on idiosyncratic household-level productivity and then aggregate up over household productivity shifters.

7 Comparison with "Expected Utility"

In contrast to our approach in this paper, by far the most common way to evaluate efficiency and aggregate welfare in these models is to use the so-called "expected utility" of the representative agent. We discuss this alternative approach, and contrast it to ours, in this section. We show that this common approach depends on an untestable cardinal assumption about utility functions that cannot be rejected by any conceivable data. This would be fine except that, other equally untestable assumptions, which are also consistent with the same data generating process, will give different answers. This stands in contrast to our measure of aggregate efficiency which depends only on ordinal properties of utility functions (i.e. it is falsifiable).

We begin by defining the common measure used in the literature, which we refer to as *average utility*.

Definition 6 (Average Utility)**.** The average utility, sometimes also called expected utility or utilitarian welfare, is given by

$$U(\boldsymbol{c},\boldsymbol{l}) = \int u_h(\boldsymbol{c}_h,\boldsymbol{l}_h)dh. \tag{10}$$

The typical approach uses $U(\boldsymbol{c},\boldsymbol{l})$ to measure aggregate welfare or efficiency. We deliberately do not refer to U as "expected utility," which is how it is sometimes referred to in the literature. The reason is because expected utility is formally defined to be a representation of an ordinal preference relation of a single agent over lotteries of allocations (see, e.g., chapter 6, of Mas-Colell et al., 1995). In this model, each household has fixed preferences u_h and there is no lottery across household tastes. This is to say, no household ever makes a choice about the parameters of their utility function, which means that we cannot elicit preferences over tastes from observed choices. Indeed, even if households were said to have preferences over their tastes, such "meta" preferences would have no testable implications.[6]

[6] This state of affairs is different to the veil-of-ignorance argument as in Harsanyi (1955). The standard veil of ignorance approach assumes all households have the same preference relations (down to risk preferences), and then asks how each household values the distribution of allocations (not utilities). Since all households have the same preferences, they all assign the same certainty equivalent value to this lottery. In this case, households have different preferences, so the veil of ignorance argument gives a different answer for each household. See Eden (2020) for a modern discussion of this point.

While (10) is a Bergson-Samuelson social welfare function, it is one choice among many possible social welfare functions, and there is nothing in the data that helps pin it down. To illustrate the fact that (10) is not pinned down by any conceivable choice data, we provide a specific example. Suppose that household preferences take the form

$$u_h(\boldsymbol{c}_h, \boldsymbol{l}_h) = \sum_r \epsilon_{hr} c_h \mathbf{1}[l_h = r],$$

where household taste parameters for choosing different options consist of two components:

$$\epsilon_{hr} = \bar{\epsilon}_h \varepsilon_{hr},$$

where the vector $\{\varepsilon_{hr}\}_r > 0$ is distributed according to the distribution of identical and independent Fréchet random variables with distribution function

$$G(\varepsilon_{hr}) = \exp[-B_r \varepsilon_{hr}^{-\theta}],$$

for some $\theta > 0$.

The location demand system associated with these preferences is the familiar and popular isoelastic one in (9). Note that this location demand system can be derived without making any assumptions about the distribution of the household-level shifter $\bar{\epsilon}_h$ and its dependence structure with the household-by-option shifters ε_{hr}. (For example, it could be that household who have a high taste for location r also have high values of $\bar{\epsilon}_h$.)

The reason that the location demand system does not depend on $\bar{\epsilon}_h$ is that h's rankings of options depend only on the relative taste shifters, $\epsilon_{hr}/\epsilon_{hr'}$, not their absolute values. Hence, since the value of $\bar{\epsilon}_h$ cancels out of this ratio, it has no bearing on the choices any household makes and therefore assumptions about it have no testable implications. In particular, the induced location demand system $\boldsymbol{L}(\boldsymbol{w})$ is the same regardless of assumptions about the distribution of the household-level $\bar{\epsilon}_h$ parameter.

The following proposition summarizes the implications for average utility.

Proposition 6 (Average Utility with Fréchet Distribution of Tastes)**.** *The average utility associated with the utility functions $\{u_h\}$ in a competitive equilibrium*

$$U = \mathbb{E}\left[\bar{\epsilon}_h \mathbb{E}\left[\max_r \{w_r \varepsilon_{hr}\} \middle| \bar{\epsilon}_h\right]\right]. \tag{11}$$

Assuming that the collection of random variables $\{\epsilon_{hr}\}$ are independent of $\bar{\epsilon}_h$, then

$$U = \Gamma\left(\frac{\theta - 1}{\theta}\right) \mathbb{E}\left[\bar{\epsilon}\right] \left(\sum_r B_r w_r^\theta\right)^{\frac{1}{\theta}}, \tag{12}$$

as long as $\theta > 1$, where Γ is the gamma function. If $\theta < 1$, then average utility, U, diverges.

The typical unstated assumption in the literature is that $\bar{\epsilon}_h = 1$ for every h (which means that it is independent of ε_{hr}). However, by inspection, Equation (11) shows that this assumption is not without loss of generality. The value of U depends on the relationship between $\bar{\epsilon}_h$ and ε_{hr}. To see one extreme case, suppose that $\bar{\epsilon}_h$ is a Dirac delta function at $(\epsilon_{h1}, \ldots, \epsilon_{hR}) = (1, \ldots, 1)$. In this limit, $U = \max_r\{w_r\}$. Hence, with this alternative assumption about $\bar{\epsilon}_h$, we arrive at very different results about the change in U due to a change in observables. Further, this alternative distribution is also consistent with the same underlying preference relations. This example highlights the fact that the value of U is not disciplined by choice data but instead depends on an untestable assumption about the joint distribution of the level of taste shifters. Choice data can only ever identify relative taste-shifters — not their levels. This is why it is not appropriate to refer to U as an expected utility function — it should instead be thought of as a social welfare function, which implicitly places weights on households based on the overall level of their ϵ's.[7] Note that this issue never arises for our measure of efficiency, because it only depends on ordinal properties of preference relations, not cardinal properties.

However, even if we impose the typical and untestable assumption that $\bar{\epsilon}_h = 1$ for every h, this measure U does not coincide with our measure of efficiency. An obvious way to see the difference is to note that U diverges to infinity if the (partial) elasticity of supply in a location to the real wage is below $\theta < 1$. Of course, in practice, these elasticities may easily be below one (i.e. a one percentage point increase in the wage raises the share of population choosing an option by less than one percent). This poses no issues for our measure of efficiency, but the average utility formulation diverges in this case.

For the isoelastic demand system, there is another justification that distinguishes (12), from any arbitrary (11). This justification is not based on invoking utilitarianism or a fictional ex-ante expected utility problem before households know their tastes. Instead,

[7] This issue is not limited to having multiplicative taste shocks. The same type of issue also arises if taste shifters are additive — in this case, what matters is the difference in taste shifters, rather than their ratios. But once again, average utility depends on the level of taste shifters. For example, if we multiply consumption and taste shifters by a different positive constant $\bar{\epsilon}_h$, we preserve the same choices. If $\bar{\epsilon}_h$ varies systematically, this changes both expected utility and changes in expected utility in response to changes in primitives.

(12) can be distinguished from other choices of (11) by the fact that it corresponds to the preferences of a *positive* representative agent. That is, a representative agent that maximizes the preferences in (12) chooses population shares and region-level consumptions that coincide with the competitive equilibrium allocations. However, whereas a positive representative agent's choices coincide with the collective actions of the underlying households, as discussed by Kirman (1992) or Tito (2016), the welfare of a positive representative agent need not have any previleged position in terms of its welfare implications for the underlying agents.

To better understand this, consider the following proposition, which shows that, to a first-order approximation, the change in average utility does not coincide with the change in $\Delta \log A$. In particular, $\Delta \log A$ obeys a version of Hulten's theorem, and hence, is first-order equal to the change in multifactor productivity as calculated by national income accountants (see, e.g., chapter 3 in OECD, 2001), the change in U does not have this property.

Proposition 7 (First-Order Changes in Average Utility with Fréchet Distribution of Tastes)**.** *Consider the environment in Section 6. Suppose that U takes the common form in* (12)*. Then, to a first-order approximation,*

$$\Delta \log U = \sum_{r \in R} \frac{L_r}{\sum_{r'} L_{r'}} \Delta \log z_r.$$

This proposition, which is easy to derive given (12), shows that $\Delta \log U$ and $\Delta \log A$ are not the same, even to a first-order approximation, unless the population share in each region coincides with the sales share of that region. One prominent example is when there is no heterogeneity in tastes (e.g. $\theta \to \infty$). However, outside of this limit, the two measures are generically different.

8 Conclusion

We illustrate how to compute aggregate efficiency in models with heterogeneous workers making discrete choices. We focus on highly stylized economies with perfect competition and no unpriced spillovers or externalities. This helps to isolate the differences between our approach and the extant literature in a simple benchmark environment. In ongoing work, we extend our analysis to cover inefficient economies, and use it to analyze distortions in location choices.

References

Anderson, S. P., A. De Palma, and J. F. Thisse (1992). *Discrete choice theory of product differentiation*.

Baqaee, D. and A. Burstein (2025a). Aggregate efficiency with heterogeneous agents. Technical report.

Baqaee, D. R. and A. Burstein (2025b, September). Efficiency costs of incomplete markets. Working Paper 34233, National Bureau of Economic Research.

Baqaee, D. R. and E. Farhi (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica 87*(4), 1155–1203.

Donald, E., M. Fukui, and Y. Miyauchi (2023). Unpacking aggregate welfare in a spatial economy. Technical report, Tech. rep.

Eden, M. (2020). Welfare analysis with heterogeneous risk preferences. *Journal of Political Economy 128*(12), 4574–4613.

Fajgelbaum, P. D. and C. Gaubert (2020). Optimal spatial policies, geography, and sorting. *The Quarterly Journal of Economics 135*(2), 959–1036.

Fajgelbaum, P. D. and C. Gaubert (2025). Optimal spatial policies. Technical report, National Bureau of Economic Research.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy 63*(4), 309–321.

Hortaçsu, A. and J. Joo (2023). Structural econometric modeling in industrial organization and quantitative marketing: theory and applications.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, 511–518.

Kim, R. and J. Vogel (2020). Trade and welfare (across local labor markets). Technical report, National Bureau of Economic Research.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of economic perspectives 6*(2), 117–136.

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic theory*, Volume 1. Oxford university press New York.

McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications 198272*.

Mongey, S. and M. E. Waugh (2024). Discrete choice, complete markets, and equilibrium. Technical report, National Bureau of Economic Research.

OECD (2001). *Measuring Productivity — OECD Manual: Measurement of Aggregate and Industry-level Productivity Growth*. Paris: OECD Publishing.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics 101*, 148–167.

Redding, S. J. and E. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics 9*(1), 21–58.

Small, K. A. and H. S. Rosen (1981). Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, 105–130.

Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312–320.

Tito, M. D. (2016). Welfare evaluation in a heterogeneous agent model: How representative is the ces representative consumer? *Economics Letters 143*, 99–102.