Aggregate Efficiency with Discrete Choice

David R. Baqaee — UCLA
Ariel Burstein — UCLA*

December 1, 2025

Abstract

This paper studies aggregate efficiency in general equilibrium models where agents with heterogeneous preferences make discrete choices between locations and occupations. We show how the conventional tools of welfare economics, like cost-benefit analysis, can be extended to these settings. Following Debreu (1951), we measure the change in aggregate efficiency by: *"the maximum reduction in total factor productivity such that it is possible to make every household at least indifferent to their status-quo allocation."* Aggregate efficiency rises if factors can be saved while keeping every household at least indifferent. We characterize our measure of efficiency in competitive economies. We show that, to a first-order approximation, its elasticity with respect to technology shocks is given by sales shares (regardless of preferences and technologies). We provide nonlinear characterizations in terms of compensated supply and demand systems. Our analysis can be viewed as a general equilibrium counterpart to the sum of compensating variations measure analyzed by Small and Rosen (1981). We also contrast our measure with the commonly used "expected utility" approach, which depends on untestable assumptions about how individual utility functions are cardinalized.

*We thank Rodrigo Adao, John Asker, Pablo Fajgelbaum, Hannes Malmberg, Will Rafey, Jon Vogel, Pierre-Olivier Weill for their comments. We are especially grateful to Natalie Bau for her insights.

1 Introduction

There has been an explosion of work on general-equilibrium economies in which agents make discrete choices over where they live and work. These models allow for heterogeneity in preferences and skills, so that different agents make different choices and face different outcomes in response to changes in the environment. Aggregating these disparate outcomes into meaningful statistics is a central task of economics. These statistics are important because they summarize overall changes in welfare and guide policy.

This paper shows how the conventional tools of welfare economics and general equilibrium theory, including cost-benefit analysis, can be extended to these environments. Before discussing our approach, we discuss shortcomings of currently popular aggregate measures in discrete choice settings. The problems we discuss with these measures motivate our search for an alternative.

i. **Chain-weighted real output**: This is how real output and consumption are measured in the national accounts. In economies with a single final good, this is simply the change in the quantity of that good. However, as is well-known, these measures can be an inappropriate gauge for welfare because they ignore amenity value (Roback, 1982). For example, if an agent moves from a high to low real wage location/job, then real output and consumption fall, even though that agent may be better off once we take into account her location preferences. Indeed, it is simple to construct Pareto-efficient examples where every agent is better off after a shock, but real output declines.[1]

ii. **Utilitarian welfare or "expected" utility**: A fundamental issue with this measure is that it is not invariant to monotone transformations of individual utility functions. This implies that replacing a utility function with a monotone transformation that generates the exact same individual choice behavior can nonetheless alter the social welfare ranking. For instance, in the spatial literature, a standard utility specification is $c\varepsilon_h$, where c is consumption and ϵ_h is h's idiosyncratic taste.[2] This is observationally equivalent to a model where tastes are normalized by each individual's average taste parameter: $\mathbb{E}(\varepsilon|h)$ (i.e., $c\varepsilon_h/\mathbb{E}(\varepsilon_h|h)$). While both specifications predict identical individual choices, average utility across agents implies a different ranking

[1] Examples of papers with discrete choice and heterogeneous preferences that analyze real output measures include, Hsieh et al. (2019) who measure misallocation of talent, Lamadon et al. (2022) who study the efficiency cost of imperfect competition (they also study effects on average utility, discussed below), Bagga et al. (2025) who study the effects of amenity shocks on labor reallocation and output, among many others.

[2] Some examples that use average utility to measure aggregate welfare in the spatial literature include Redding (2016), Caliendo et al. (2019), and Allen and Arkolakis (2022).

of social allocations. Crucially, there is no canonical basis for selection here: normalizing average taste intensity to unity is just as plausible as leaving it unscaled. Since the data cannot distinguish between these scaling methods, the resulting aggregate welfare measure relies on untestable assumptions regarding cardinalization.[3]

iii. **Sum of compensating variations**: This measure is also called Kaldor-Hicks efficiency or cost-benefit efficiency, and has been analyzed by Small and Rosen (1981) for settings with discrete choice. It measures the amount of money left-over after winners compensate the losers holding new equilibrium prices and wages constant. The pioneers in the discrete choice literature, like McFadden (1981) and Anderson et al. (1992), emphasize that consumer welfare in discrete choice models should be measured using compensating or equivalent variations of the individual agents, as in standard consumer theory.[4] While this measure does not have the issues of the previous two, it can nevertheless be unreliable in general equilibrium. In particular, the implied compensations may be infeasible. For example, we show that Kaldor-Hicks efficiency can rise in response to pure transfers that move the economy along the Pareto-efficient frontier. In this sense, the sum of compensating variations is not a pure measure of efficiency.

Our approach, which side-steps these issues, is a reformulation of the third approach. We follow in the footsteps of Harberger (1971) and Small and Rosen (1981) in using willingness-to-pay as a basis for constructing a measure of aggregate efficiency. However, instead of adding up willingness-to-pay at constant prices, we use a definition in terms of quantities. This definition of efficiency, proposed by Debreu (1951) and, more recently extended by Baqaee and Burstein (2025) asks: "*what is the maximum reduction in total factor productivity such that it is possible to make every household at least indifferent to their status-quo allocation?*" If we can save x% of factor endowments and keep every household at least indifferent, then we say that efficiency has increased by x%.

Kaldor-Hicks measures efficiency gains by the amount of aggregate income that could be taken away — holding prices and wages fixed — while leaving everyone as well-off as under the status-quo. Our definition is similar, but it measures efficiency gains in terms of

[3]See Section 6.1 for a discussion and worked out example. Although sometimes called "expected utility," measures of average utility are not formally expected utility functions. This is because expected utility functions are representations of ordinal preference relations on the space of lotteries — not an expectation of utility values across households with different preferences.

[4]While this approach is common in the industrial organization literature, it is less common in general equilibrium spatial and occupational choice models. One exception is Kim and Vogel (2020), who study a model where workers choose among a discrete number of sectors including non-employment. They show that the elasticity of the sum of compensating variations with respect to wages is given by income share of each sector.

primary factors rather than dollars: it asks how much of those factors could be removed — holding technologies fixed — while leaving everyone indifferent to the status-quo.

This measure has several useful features. First, the answer has interpretable units expressed in terms of total factor productivity (or equivalently, in units of every good, since scaling total factor productivity also scales the consumption possibility set). Second, and unlike real output, it respects the Pareto principle — efficiency rises only if every agent can be made better off. Third, and unlike "expected" utility, the question is about observables with a falsifiable answer. That is, it is invariant to monotone transformations of utility and does not introduce other free parameters like Pareto weights. Hence, models that are observationally equivalent in terms of physical allocations, will assign the same number to the change in aggregate efficiency under our definition. Fourth, unlike the sum of compensating variations, the question is stated and the answer is measured in terms of exogenous variables, rather than endogenous prices and wages. This implies, for example, that aggregate efficiency is unaffected by pure redistributions that move the economy along the Pareto frontier. Fifth, it does not take a stance on distributional considerations or embed normative judgements about interpersonal utility comparisons. Finally, it is relatively easy to communicate its meaning to non-experts and policy-makers.

We study this question in perfectly competitive but otherwise fairly flexible general equilibrium economies. Focusing on a perfectly competitive, and hence Pareto-efficient, benchmark helps clarify the differences between our approach and standard practice in the literature. We abstract from distortions and externalities discussed by, for example, Fajgelbaum and Gaubert (2020), but 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 and define perfectly competitive equilibrium 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. In Section 3, we define our notion of aggregate efficiency and show how it can be characterized using compensated (or Hicksian) supply and demand functions. In this context, compensated supply functions map relative wages and prices to population shares across locations or occupations, holding each household on a fixed indifference curve. We also show how to compute compensated supply functions using simulation methods.

In Section 4, we provide some analytical (as opposed to simulation-based) characterizations of our measure. First, we show that our measure of aggregate efficiency, to a first-order approximation, obeys Hulten's theorem. That is, the elasticity of aggregate efficiency to a productivity shock to producer i is equal to the sales of i divided by total income. In particular, one does not need to know anything about the underlying preferences or technologies. Moreover, one does not need to take into account changes in behavior in response to the shock (to a first order). Intuitively, any changes in real wages experienced by agents that switch their choice in response to a shock are exactly offset by changes in the amenity value of their choices. We show that real output, in contrast, does not obey Hulten (1978) and hence differs from aggregate efficiency even to a first-order (despite the fact that the model is efficient).

Whereas our measure of efficiency differs from real output, it does coincide, 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.

In this section, we also point out a noteworthy special case: agents only care about a common consumption good (no amenities) and vary in their skills, rather than tastes. Economies satisfying these assumptions are common in the literature on skill-based occupational choice driven by comparative advantage. In this special case, our measure of aggregate efficiency coincides with multi-factor productivity growth, not only to a first-order but globally. In fact, in this case, aggregate efficiency also coincides with real output.

In Section 5, we push our analytical characterization beyond the first-order and investigate nonlinearities. We provide a second-order approximation of changes in aggregate efficiency. We show that the first-order approximation understates positive shocks and overstates negative shocks if compensated sales shares rise for producers with positive shocks. The reverse is true if compensated sales shares fall in response to positive shocks. We express the elasticities of compensated sales shares in terms observables in the initial equilibrium: income and expenditure shares, and price elasticities of uncompensated (market-level) supply and demand systems. We also provide a tractable special case where we can exactly characterize aggregate efficiency, beyond second-order, and without the use of simulation methods. This is the case when all agents consume a common freely-traded consumption good and the uncompensated supply system has symmetric cross-partials (e.g. population choices obey a logit supply system). We also provide some analytical examples.

In Section 6, we compare our approach to the sum of compensating variations (Kaldor-Hicks efficiency) studied by, for example, Small and Rosen (1981). We emphasize that when we refer to Kaldor-Hicks we do not refer to a partial equilibrium exercise where some equilibrium variables or choices are kept constant at their initial values. That is, we always measure Kaldor-Hicks efficiency using prices and wages in general equilibrium.[5] We show that our measure does not generally coincide with Kaldor-Hicks efficiency by providing an example where Kaldor-Hicks efficiency detects increases in efficiency in response to pure transfers as the allocation moves along the Pareto-frontier.

In Section 6, we also compare our approach to the utilitarian/"expected" utility approach. We show that the utilitarian approach typically does not coincide with our approach, even to a first-order. As discussed above, we also show that different arbitrary but equally plausible monotone transformations of the individual utility functions result in different utilitarian social welfare functions with different implications for optimal policy even in Pareto-efficient models. We do discuss one special cardinalization of individual functions, where under some additional strong assumptions, a utilitarian calculation coincides with our measure of aggregate efficiency.[6]

For a detailed analysis of aggregate welfare using social welfare functions in spatial models, see Donald et al. (2023). 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") utility function. Instead, we define aggregate efficiency in terms of resource savings after compensating transfers.

This difference is important because utilitarian welfare or "expected" utility conflates efficiency considerations with distributional concerns (see the discussion by Fajgelbaum and Gaubert, 2025). In particular, policymakers looking to maximize average utility have incentives to intervene in Pareto-efficient allocations for purely distributive motives.[7] In

[5]Contrast this with traditional partial equilibrium cost-benefit analysis, which ignores the fact that equilibrium variables can change. (See the discussion by Redding, 2025).

[6]This relates to the a well-known property of discrete choice models where consumers choose among a discrete set of options given fixed prices and incomes. In that literature, it is known that if the indirect utility function is quasi-linear in income, then exists a closed-form formula, known as the "social surplus function," that can be used to calculate the sum of compensating variations (see Small and Rosen, 1981). 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 in the industrial organization literature ignore individual 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, even if the quasi-linear justification for this approach does not hold. Dubé et al. (2025) analyze the welfare properties of this expected social surplus function for a broad class of demand systems.

[7]Mongey and Waugh (2024) provide an alternative interpretation where average utility is the expected utility of individuals who are ex-ante identical (before their taste shocks are known) and lack access to insurance markets for their idiosyncratic tastes. However, if such an ex-ante state does not exist, then the results of this type of analysis also rest on an untestable cardinalizing assumption. That is, if households cannot make choices in the ex-ante stage that reveal their risk preferences about how they rank one set of taste parameters against another, then choice data is only determined up to a monotone transformation and

contrast to such a measure, our measure of efficiency is maximized in Pareto-efficient allocations. That is, if a policy yields a higher value of aggregate efficiency, then it must be that this policy can make all households better off.

2 Environment and Equilibrium

In this section we define the environment and equilibrium. We consider perfectly competitive economies without externalities (e.g. positive or negative spillovers). There is a set of commodities, N, and a set of primary factors R. As in Arrow-Debreu, commodities may be indexed by region (e.g. haircuts in LA and New York can be treated as different commodities).[8]

Primary factors can be arbitrarily finely differentiated in the set R. For example, the set R can distinguish factors by region, occupation, skill-level, and type (e.g. high-skilled accountants in LA and commercial land in New York are different factors). Agents have different ability to provide the different factor services. For example, some agents (low-skilled workers) can only supply low-skilled labor, or landowners in one region can only supply land to firms in their region but not outside. We refer to the discrete choice $l_h \in R$ as the *location* choice of h with the understanding that it may index location in space as well as occupation, industry, and so on.

Households' problem. There is a set H of agents, with a unit mass. Each agent $h \in H$ consumes a scalar-valued (homothetic) bundle c_h (e.g. of goods and services) and a location $l_h \in R$ (e.g. being a barber in Los Angeles or a waiter in New York, etc.) Since goods may be indexed by location, the bundle c_h may depend on the choice of location (e.g. barbers in Los Angeles consume a different bundle of goods than waiters in New York). Agent h has preferences $\succeq_h$ over the bundle of consumption and location choice represented by a utility function $u_h(c_h, l_h)$.

Household h's preference are ordinally additively separable between location choices and consumption. This means that the utility function can be written as

$$u_h(c_h, l_h) = f_h(g(c_h) + \epsilon_{hl_h}),$$

the ex-ante expected utiltiy function can not be recovered from the data. See also the discussion of this issue by Davis and Gregory (2021).

[8]Similarly, if there are trade costs, then traded goods are indexed by both origin and destination. For example, oranges produced in California and consumed in New York are different to oranges produced in California and consumed in Illinois.

where g is a strictly increasing function, ϵ_{hl_h} is a taste parameter for household h's preferences for location l_h, and f_h is any strictly increasing (potentially household-specific) function.[9]

The household maximizes utility, choosing consumption c_h and location l_h. The household's budget constraint is that consumption expenditures are less than or equal to income plus transfers:

$$\sum_{r \in R} p_r c_h \mathbf{1}[l_h = r] \leq \sum_{r \in R} a_{hr} w_r \mathbf{1}[l_h = r] + T_h,$$

where p_r is the price of the consumption bundle in region r, a_{hr} is agent h's efficiency units of labor given choice r, w_r is the wage per efficiency unit in that location, and T_h is a lump-sum transfer. We also anticipate that firms earn zero profits in equilibrium. Recall again that the choice of location encompasses both spatial location and choice of occupation.

Denote the efficiency units of labor in location r by

$$L_r = \int a_{hr} \mathbf{1}\left[l_h = r\right] dh.$$

We call the function $\boldsymbol{L}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{T})$ — mapping vectors of prices, wages, and lump-sum transfers into the share of households in each location — the aggregate *labor supply function*. Note each h chooses the location r that yields the highest $g((a_{hr} w_r + T_h)/p_r) + \epsilon_{hr}$.

Let $\boldsymbol{\chi}$ denote the vector of final consumption shares relative to GDP, so

$$\chi_r = \frac{w_r L_r + \int T_h \mathbf{1}[l_h = r] dh}{\sum_{r'} w_{r'} L_{r'}}.$$

We call the function $\boldsymbol{\chi}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{T})$ — mapping vectors of prices, wages, and lump-sum transfers into share of total spending by households in each location — the *final demand* function. We assume that, in the initial decentralized equilibrium, there are no lump-sum transfers, $T_h = 0$, in which case χ_r is pinned down by L_r and wages.

Producers' problem. The set of commodities is N. This includes consumption bundles in different locations, as well as other goods used as intermediate inputs. Each good $i \in N$

[9] Since utility is only pinned down up to monotone transformations, the function f_h has no observable implications and can vary at the level of each h in arbitrary ways. However, the shape of $g(c_h)$ has testable implications, since it controls the way income and substitution effects interact with each other. In particular, $g(\cdot)$ affects how households switch choices in response to lump-sum transfers. For example, if $g(\cdot)$ is an affine function, then households do not switch their choices if we add a constant amount to their consumption in every location. By contrast, if $g(\cdot)$ is log, then households do not switch their choices if we multiply their consumption by a constant in every location. See Appendix A for more information.

is produced by perfectly competitive firms that maximize profits

$$p_i y_i - \sum_{j \in N} p_j x_{ij} - \sum_{r \in R} w_r L_{ir},$$

subject to a constant-returns technology

$$y_i = z_i F_i \left(\{x_{ij}\}_{j \in N}, \{Z L_{ir}\}_{r \in R} \right), \tag{1}$$

where z_i is a Hicks-neutral productivity shifter for producer i, F_i is a constant returns production technology, x_{ij} are intermediate inputs from $j \in N$, L_{ir} are factor services from r. The scalar Z is an aggregate factor-augmenting productivity (or total factor productivity) shifter that uniformly scales the production possibility set. We normalize $Z = 1$ in the initial equilibrium.

Resource Constraints. The resource constraint for good $i \in N$ is that consumption and intermediate input usage is less than production. Without loss of generality, we split goods into pure consumption goods and pure intermediates. That is, if $i \in R$, then i is a consumption good and $x_{ji} = 0$. On the other hand, if $i \notin R$, then i is an intermediate good and is not directly consumed by any household. Hence the resource constraint for good i is

$$\int c_h \mathbf{1}[l_h = i] dh + \sum_j x_{ji} \leq y_i. \tag{2}$$

We say that there is a *common consumption good* when the price of the consumption good in every location is the same (e.g. the consumption aggregator in each location is the same as every other location).

The resource constraint for factor r requires that total factor demand from producers is less than total factor supply from households:

$$\sum_{i \in N} L_{ir} \leq L_r. \tag{3}$$

Finally, lump-sum transfers add up to zero:

$$\int T_h dh = 0. \tag{4}$$

Definition 1 (Equilibrium with Discrete Choice)**.** An equilibrium is a collection of consumptions, c_h, location choices, l_h, outputs, y_i, intermediate input choices, $\{x_{ij}, L_{ir}\}$, prices, p_i and wages, w_r, such that each agent chooses consumption and location to maximize utility subject to their budget constraint, producers maximize profits subject to technology taking prices as given, transfers add up to zero, and all resource constraints are satisfied.

Solving for Equilbrium. To solve for the equilibrium, we introduce some useful accounting variables. Let λ be the vector of Domar (1961) weights. The ith component is the sales of i divided by total income:

$$\lambda_i = \frac{p_i y_i}{\sum_r w_r L_r}.$$

Let Ω be the $N \times N$ input-output matrix, whose ijth element is:

$$\Omega_{ij} = \frac{p_i x_{ij}}{p_i y_i}.$$

Since consumption bundles are treated like any other commodity in the set N, we use Ω_{ri} to denote the budget share of the consumption bundle of location r on commodity i (e.g. the budget share of consumption in New York spent on oranges from Florida). With some abuse of notation, when $i \in N$ is a good and $r \in R$ is a location, we define Ω_{ir} to be i's expenditures on labor of type r relative to its sales:

$$\Omega_{ir} = \frac{w_r L_{ir}}{p_i y_i}.$$

(e.g. the share of sales of Florida oranges spent on agriculture labor in Florida).

Using this notation, we can now describe the equilibrium conditions. Perfect competition and cost-minimization by producer i implies that p_i is equal to its marginal cost:

$$p_i = z_i^{-1} \mathrm{mc}_i(\boldsymbol{p}, Z^{-1}\boldsymbol{w}), \tag{5}$$

which depends on input prices, factor wages, and productivity shifters z_i and Z.

Goods market clearing for good i requires that total sales of good i equal total spending on good i by households and other producers:

$$\lambda_i = \sum_{r \in R} \chi_r \Omega_{ri} + \sum_{j \in N} \lambda_j \Omega_{ji}. \tag{6}$$

Factor market clearing for labor in location r requires that total income earned by workers in location r equals total spending by producers:

$$\lambda_r \equiv \frac{w_r L_r}{\sum_{r'} w_{r'} L_{r'}} = \sum_{j \in N} \lambda_j \Omega_{jr}. \tag{7}$$

In an equilibrium without lump-sum transfers, the share of spending in each location is equal to the share of income earned in that region:

$$\chi_r = \lambda_r. \tag{8}$$

Hence, given the supply function $L_r(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{T})$, one can solve for equilibrium prices and quantities using (5), (6), (7), and (8).

The example below provides a concrete illustration of the structure of the model using Cobb-Douglas and logit functional forms.

Example 1 (Illustration of Equilibrium). Every good is produced using a Cobb-Douglas production technology, so (5) implies that the price of each good can be written as

$$p_i = z_i^{-1} \prod_{j \in N} p_j^{\Omega_{ij}} \prod_{r \in R} (w_r / Z)^{\Omega_{ir}},$$

where Ω_{ij} and Ω_{ir} are expenditure shares of i on j and r respectively (Cobb-Douglas implies that these expenditure shares are constant). Goods market clearing, (6), implies that

$$\lambda_i = \sum_r \chi_r \Psi_{ri},$$

where $\Psi = (I - \Omega)^{-1}$ is the Leontief inverse (which, again, is constant due to the Cobb-Douglas assumption). Finally, factor market clearing, (7), implies that

$$\lambda_r = \frac{w_r L_r}{\sum_{r''} w_{r''} L_{r''}} = \sum_{j \in N} \sum_{r' \in R} \chi_{r'} \Psi_{r'j} \Omega_{jr} = \sum_{r' \in R} \chi_{r'} \Psi_{r'r}$$

where the final equality uses the identity that $\Psi\Omega = \Psi - I$. This equation states that the income earned by factor r, as a share of total income, must equal the dollar-weighted average factor content of final consumption.

Suppose that all consumers consume the same consumption good, so that the factor content $\Psi_{r'r}$ is the same for every r'. The previous equation simplifies to

$$\lambda_r = \frac{w_r L_r}{\sum_{r''} w_{r''} L_{r''}} = \Psi_{0r}, \tag{9}$$

where 0 is the index for the common consumption good and the right-hand side is a constant, depending only on the Cobb-Douglas share parameters.

Suppose skills are homogeneous, $a_{hr} = 1$, and household preferences take the form $u_h(c_h, l_h) = f_h(c_h + \epsilon_{hr}\mathbf{1}[l_h = r])$ where ϵ_{hr} are drawn from type I extreme value distribution and f_h is any strictly increasing function. We get the standard McFadden (1973) result that the labor supply function is a logit function:

$$L_r(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{T}) = \frac{\exp(\theta w_r / p_0 + B_r)}{\sum_{r'} \exp(\theta w'_r / p_0 + B'_r)}, \tag{10}$$

where p_0 is the price of the final consumption good and θ and B_r are parameters of the distribution of ϵ_{hr}. Note that lump-sum transfers do not affect location choices because $g(c_h)$ is linear and there is only a single consumption good (i.e. transfers T_h cancel out when ranking $g((a_{hr}w_r + T_h)/p_r) + \epsilon_{hr}$ across locations). Given this functional form, equation (9) can be rewritten as

$$\frac{w_r \exp(\theta w_r / p_0 + B_r)}{\sum_{r'} w_{r'} \exp(\theta w'_r / p_0 + B'_r)} = \Psi_{0r}$$

where the consumer price index satisfies

$$p_0 = Z^{-1} \prod_j z_j^{-\Psi_{0j}} \prod_r w_r^{\Psi_{0r}}.$$

This pins down all equilibrium prices, up to the choice of numeraire, which can then be used to pin down quantities.

If we strip out intermediates from the model, then we get an even simpler model where each location represents the output of some occupation, and occupation outputs are combined using a Cobb-Douglas aggregator with parameters Ω_{0r} to produce the consumption good (i.e. $\Omega_{ij} = 0$ for $i \neq 0$ and $j \in N$)). In this case, equilibrium wages and prices satisfy:

$$\frac{w_r \exp(\theta w_r / p_0 + B_r)}{\sum_{r'} w_{r'} \exp(\theta w'_r / p_0 + B'_r)} = \Omega_{0r}$$

and

$$p_0 = Z^{-1} \prod_r (w_r / z_r)^{\Omega_{0r}}.$$

---

3 Defining and Characterizing Aggregate Efficiency

In this section, we define aggregate efficiency. For convenience, index technologies $\boldsymbol{z}(t)$ by a scalar t. We study changes in aggregate efficiency as t changes, assuming that $t = 0$ corresponds to an initial equilibrium allocation called the *status-quo*. For each equilibrium price and quantity, X, we write $X(t)$.

3.1 Definition

Define the set of feasible allocations at t, given productivity shifters $\boldsymbol{z}$ to be

$$\mathcal{C}(t, Z) \equiv \{\{\boldsymbol{c}_h, l_h\}_{h\in H} \text{ is feasible given productivities } \boldsymbol{z}(t) \text{ and factor-augmenting productivity } Z\}.$$

We define aggregate efficiency following Debreu (1951).

Definition 2 (Aggregate efficiency)**.** Aggregate efficiency at t relative to $t = 0$ is

$$A(t) = \max\left\{Z^{-1} : \{c_h, l_h\}_{h\in H} \in \mathcal{C}(t, Z) \text{ and } (c_h, l_h) \succeq_h (c_h(0), l_h(0)) \text{ for every } h\right\}. \quad (11)$$

In words, $A(t)$ is the maximum contraction of total factor productivity, Z, at t, such that we can feasibly keep every agent at least as well off as in the status-quo. Intuitively, if $A(t) = 1.01$, then this means that we can shrink the productivity of factors by roughly 1% (more precisely, $1 - 1.01^{-1}$) and still keep every household indifferent to the initial equilibrium. Equivalently, since Z uniformly scales the production possibility set, this means we can ensure indifference with roughly 1% of every consumption good left over.

3.2 Exact Characterization

Characterizing $A(t)$ by solving (11) is challenging. It requires jointly picking locations and consumption for all households such that, once we divide factor productivity by A, every household is indifferent to the status-quo and A is as big as possible. Solving this problem by brute force is impractical because it involves solving a very large general-equilibrium combinatorial problem where the choice variable is a discrete vector $\{l_h\}_h$ and continuous vector $\{c_h\}_h$ both of length H belonging to the feasible set.

To solve this problem, we instead show that the allocation corresponding to $A(t)$ is supported by a decentralized equilibrium with appropriate lump-sum transfers. We refer to this as the *compensated* equilibrium (Baqaee and Burstein, 2025, use similar terminology). In the compensated equilibrium, every household is kept indifferent to the status-quo through the adjustment of lump-sum transfers and aggregate factor productivity. The compensated equilibrium provides a much more efficient way to solve for $A(t)$ nonlinearly, and it allows us to tie the local (marginal) properties of $A(t)$ to observed expenditure shares and price elasticities of supply and demand curves in the decentralized equilibrium.

3.2.1 Expenditure Function and Compensated Choices

To do this, we begin by defining the expenditure function of each consumer.

Definition 3 (Expenditure function)**.** Define $e_h(\boldsymbol{p}, \boldsymbol{w}, u_h^0)$ to be the (net) expenditure function for agent h given prices $\boldsymbol{p}$, wages $\boldsymbol{w}$, and utility level u_h^0:

$$e_h(\boldsymbol{p}, \boldsymbol{w}, u_h^0) = \min\{T_h : u_h(c_h, l_h) \geq u_h^0, \text{ and } \sum_r p_r c_h \mathbf{1}[l_h = r] \leq \sum_r a_{hr} w_r \mathbf{1}[l_h = r] + T_h\}.$$

We call the location choice $l_h^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, u_h^0)$ associated with this optimization problem *the compensated choice* of agent h.

In words, $e_h(\boldsymbol{p}, \boldsymbol{w}, u_h^0)$ is the minimum lump-sum transfer agent h requires to be made at least indifferent to some reference utility level u_h^0, given prices and wages. The location choice that the agent makes, given that transfer, is what we call the compensated choice. This definition parallels the classical definition of expenditure functions and compensated demand in traditional consumer theory.[10]

To characterize the compensated choice of each household, we first define the consumption-equivalent variation.

Definition 4 (Consumption-equivalent variation)**.** Define $\bar{c}_{hl_h}$ as the solution to

$$g(\bar{c}_{hl_h}) + \epsilon_{hl_h} = g(c_h^0) + \epsilon_{hl_h^0},$$

where c_h^0 is the consumption and l_h^0 is the location of h in the status-quo. Hence, $\bar{c}_{hl_h}$ is the consumption agent h must have in location l_h to be indifferent to the status-quo.

Given the consumption-equivalent variation, the following proposition shows how to compute the compensated choices and the expenditure function at the household level.

[10] See also Small and Rosen (1981) for a related analysis. Whereas they consider discrete consumption choices, given fixed income, here the level of income depends also on the discrete choice (via the wage).

Proposition 1 (Compensated Choices and Expenditure Function)**.** *The compensated choice* $l_h^{comp}(\boldsymbol{p}, \boldsymbol{w}, u_h^0)$ *satisfies*

$$l_h^{comp}(\boldsymbol{p}, \boldsymbol{w}, u_h^0) \in \arg\max_{l \in R} \left[\sum_r \left[a_{hr} w_r - p_r \bar{c}_{hr} \right] \mathbf{1}\left[l = r \right] \right].$$

The expenditure function, or transfer needed to ensure indifference to u_h^0*, is*

$$e_h(\boldsymbol{p}, \boldsymbol{w}, u_h^0) = p_{l_h^{comp}} \bar{c}_{h l_h^{comp}} - a_{hr} w_{l_h^{comp}}.$$

In words, $[a_{hr} w_r - p_r \bar{c}_{hr}]$ is the surplus, in dollars, household h receives from being sent to location r. The compensated choice maximizes this surplus, and the expenditure function is equal to the negative of the surplus. This proposition gives a straightforward way to calculate the compensated choice and the compensating transfer for each agent given a vector of prices $\boldsymbol{p}$ and wages $\boldsymbol{w}$.

Given household-level location choices and expenditure functions from Proposition 1, we can aggregate to get location-level variables. The compensated labor supply function, the efficiency of units of labor in location r given compensating transfers, is:

$$L_r^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{u}^0) = \int a_{hr} \mathbf{1}[l_h^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, u_h^0) = r] dh.$$

Similarly, the compensated spending by households that choose location r is the sum of their labor income and net transfer payments:

$$E_r^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{u}^0) = w_r L_r^{\text{comp}} + \int e_h(\boldsymbol{p}, \boldsymbol{w}, u_h^0) \mathbf{1}[l_h^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, u_h^0) = r] dh,$$

where $l_h^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, u_h^0)$ and $e_h(\boldsymbol{p}, \boldsymbol{w}, u_h^0)$ are given by Proposition 1. We can also denote the compensated share of total spending by households in location r to be

$$\chi_r^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{u}^0) = \frac{E_r^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{u}^0)}{\sum_{r'} E_{r'}^{\text{comp}}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{u}^0)}.$$

Hence, Proposition 1 provides a way to evaluate the compensated share function $\boldsymbol{L}^{\text{comp}}$ and expenditures $\boldsymbol{E}^{\text{comp}}$ given any arbitrary vector of wages, prices, and utility levels.

3.2.2 Calculating $A(t)$ using Compensated Choices and Spending

The next result uses the compensated regional location choices and expenditures to solve for $A(t)$.

Theorem 1 (Aggregate Efficiency via Compensated Equilibrium)**.** *Aggregate efficiency satisfies*

$$A(t) = \frac{\sum_r w_r L_r^{comp}(\boldsymbol{p}, \boldsymbol{w}/A, \boldsymbol{u}^0)}{\sum_r E_r^{comp}(\boldsymbol{p}, \boldsymbol{w}/A, \boldsymbol{u}^0)}, \tag{12}$$

The vector of prices $\boldsymbol{p}$ *equal marginal costs,* (5)*, given wages* $\boldsymbol{w}$ *and productivities* $\boldsymbol{z}$*. The vector of wages* $\boldsymbol{w}$ *satisfy market clearing conditions* (6) *and* (7)*, given compensated location choices* $\boldsymbol{L}^{comp}$ *and expenditure shares* $\boldsymbol{\chi}^{comp}$*. The wages, prices, location choces, and quantities that satisfy these conditions are not the same as the ones in the decentralized equilibrium. We call them variables in the* compensated *equilibrium instead.*

Solving for $A(t)$ using Theorem 1 is relatively straightforward. To see the logic, consider the following iterative procedure. First, conjecture a vector of wages in the compensated equilibrium, $\boldsymbol{w}$, and aggregate efficiency A. Use the wages and the productivities, $\boldsymbol{z}(t)$, to solve for prices by setting prices equal to marginal costs, as in (5). Then apply Proposition 1 to get compensated choices $\boldsymbol{L}^{\text{comp}}$ and spending $\boldsymbol{E}^{\text{comp}}$ given prices, wages, and conjectured aggregate efficiency. Given $\boldsymbol{L}^{\text{comp}}$ and $\boldsymbol{E}^{\text{comp}}$, use equations (6) and (7) to update wages and use equation (12) to update aggregate efficiency. In a compensated equilibrium, these new wages and measures of aggregate efficiency coincide with the initial conjectures.

3.3 Single-Good Example

We illustrate Theorem 1 using a simple example. Suppose there is a single consumption good produced linearly from labor with productivity $z_r(t)$ in location r. Assume households skill is homogeneous: $a_{hr} = 1$ for every $h \in H$ and $r \in R$. The production technology in each region is

$$y_r = z_r L_r = z_r \int \mathbf{1}[l_h = r] dh,$$

where L_r is the share of households in location r. Consumption goods produced in each region are perfectly substitutable and freely traded. The aggregate resource constraint is

$$\int c_h dh \leq \sum_r z_r(t) L_r.$$

The feasible set requires that total consumption equal total production:

$$\mathcal{C}(t) = \left\{ \{\boldsymbol{c}_h, \boldsymbol{l}_h\} : \int c_h dh \leq \sum_r z_r \int [l_h = r] dh \right\}.$$

Applying Theorem 1 is simple in this economy since the real wage in the compensated equilibrium must be z_r. Hence, (12) is

$$A(t) = \frac{\sum_r \int z_r(t) \mathbf{1}[l_h^{\text{comp}} = r] dh}{\sum_r \int \bar{c}_{hr} \mathbf{1}[l_h^{\text{comp}} = r] dh}, \tag{13}$$

where, by Proposition 1, the compensated choices satisfy:

$$l_h^{\text{comp}} \in \arg\max_{l \in R} \left[\sum_r \left[z_r / A - \bar{c}_{hr} \right] \mathbf{1} \left[l = r \right] \right].$$

This is a simple numerical problem to solve agent-by-agent. However, to get a more explicit solution, we can further assume that $g(c)$ is linear. In this case, by Definition 4

$$\bar{c}_{hl_h} = z_{l_h^0}(0) + \epsilon_{hl_h^0} - \epsilon_{hl_h}.$$

Substituting this into (13) and rearranging yields

$$\sum_r \int \left[z_r(0) + \epsilon_{hr} \right] \mathbf{1} \left[l_h^0 = r \right] = \sum_r \int \frac{z_r(t)}{A(t)} \mathbf{1}[l_h^{\text{comp}} = r] dh + \epsilon_{hl_h}. \tag{14}$$

Substituting $\bar{c}_{hl_h}$ into the expression for l^{comp} in Proposition 1 we also see that

$$l_h^{\text{comp}} \in \arg\max_l \sum_r [z_r / A + \epsilon_{hr}] \mathbf{1}[l = r] = \arg\max_l u_h(z_l / A, l).$$

Hence, the compensated location choice is the same as the uncompensated location choice given the vector of real wages $\boldsymbol{z}/A$. Hence, we can rewrite (14) as

$$\mathbb{E} \left[\max_r z_r(0) + \epsilon_{hr} \right] = \mathbb{E} \left[\max_r \frac{z_r(t)}{A(t)} + \epsilon_{hr} \right].$$

where the expectation averages over all households. Under the common type I extreme value distribution, we get that

$$\sum_r \exp(\theta z_r(0) + B_r) = \sum_r \exp(\theta z_r(t) / A(t) + B_r). \tag{15}$$

This is a single nonlinear equation that can be solved to yield $A(t)$.

Dangerous leap. A serendipitous fact about (15) is that $A(t)$ coincides with computing utilitarian welfare assuming a particular cardinalization of the ordinal preference relations. In particular, suppose we use a particular (arbitrary) cardinalization of the utility function: $u_h(c_h, l_h) = f_h(g(c_h) + \epsilon_{hl_h}) = c_h + \epsilon_{hl_h}$, and we define $W(\boldsymbol{u}) = \int u_h dh$ to be utilitarian social welfare under this cardinalization. Then, (15) implies that $A(t)$ satisfies:

$$\max_{\boldsymbol{l}} W(u_h(\boldsymbol{z}(0))) = \max_{\boldsymbol{l}} W(u_h(\boldsymbol{z}(t)/A(t))).$$

This is the change in productivity a utilitarian planner needs to be indifferent to the status-quo. This strategy, which works when $g(c)$ is a linear function and there is a single consumption good, breaks down whenever $g(c)$ is nonlinear or if consumption goods in different locations have different prices. See Section 6 for more details.[11]

4 Analytical Results for Aggregate Efficiency

In this section, we provide analytical results for aggregate efficiency in terms of elasticities and integrals of compensated quantities. We compare $A(t)$ with real output changes, since real output is a commonly used metric for aggregate efficiency and it is also instructive.

For any variable X, we use $X(t)$ and $X^{\text{comp}}(t)$ to refer to its value in the decentralized equilibrium and in compensated equilibrium (the compensated equilibrium is described in Theorem 1).

4.1 Sufficient Statistics for Aggregate Efficiency

In this section, we provide hat algebra-style characterizations of aggregate efficiency in terms of expenditure shares and price elasticities. We begin with the following result showing that aggregate efficiency is the sum of microeconomic productivity shocks, weighted by Domar weights in the compensated equilibrium.

Theorem 2 (Efficiency as Area Under Compensated Sales Shares)**.** *The change in aggregate efficiency satisfies*

$$\log A(t) = \int_0^t \sum_i \lambda_i^{comp}(s) \frac{d \log z_i}{ds} ds,$$

[11] This result is a counterpart to the quasi-linear indirect utility case discussed by Small and Rosen (1981).

where $\lambda_i^{comp}(s)$ is the compensated sales of i divided by GDP given productivities $\boldsymbol{z}(s)$.

Using Theorem 2 requires knowledge of sales shares, $\lambda_i^{\text{comp}}(s)$, in the compensated equilibrium. A useful fact about variables in the compensated equilibrium is that at the status-quo, compensated variables are equal to decentralized variables. This is because the decentralized equilibrium is Pareto efficient (the first welfare theorem holds), and hence, $A(0) = 1$ and the decentralized equilibrium supports the compensated allocation with no transfers needed. This allows us to establish the following useful corollary.

Corollary 1 (First-Order Approximation)**.** *To a first-order approximation, the change in aggregate efficiency satisfies*

$$\log A \approx \sum_i \lambda_i(0) \Delta \log z_i,$$

where $\lambda_i(0)$ is the sales of i divided by GDP at the status-quo.

Corollary 1 is a generalization of Hulten's theorem to economies with discrete choice.[12] Using Theorem 2 beyond first-order approximations requires knowledge of sales shares, $\lambda_i^{\text{comp}}(t)$, in the compensated equilibrium away from status-quo $t > 0$. This is easy to do, using standard methods, if we know the compensated location choices, L^{comp} and compensated final demand χ_r^{comp}. Given these, computing $\lambda^{\text{comp}}(s)$ follows from the results in Baqaee and Farhi (2019).

For example, assume that all production and consumption functions are CES, with the *i*th producer in the input-output matrix Ω corresponding to a CES aggregator with elasticity of substitution θ_i. Then the following proposition determines sales shares in both the compensated and uncompensated equilibrium.[13]

Proposition 2 (Prices and Expenditures in Equilibrium)**.** *The following differential equations*

[12] Corollary 1 is obtained by differentiating the expression in Theorem 2 with respect to t and evaluating at $t = 0$, combined with the observation that compensated variables coincide with decentralized variables in the status-quo.

[13] This structure is flexible enough to capture any nested-CES production and consumption functions through relabeling. One can extend to non-nested CES along similar lines to Baqaee and Farhi (2019).

determine the evolution of sales shares in the decentralized equilibrium:

$$d\log\Omega_{ij} = (1-\theta_i)\left[d\log p_j - d\log p_i + d\log z_i\right], \tag{16}$$

$$d\log p_i = -d\log z_i + \sum_j \Omega_{ij} d\log p_j + \sum_f \Omega_{if} d\log w_r, \tag{17}$$

$$d\log\lambda_r = d\log w_r + d\log L_r - \sum_{r'} \lambda_{r'}\left[d\log w_{r'} + d\log L_{r'}\right] \text{ for } r \in R \tag{18}$$

$$d\log\lambda_i = \lambda_i^{-1}\left(\sum_r d\chi_r\Omega_{ri} + \sum_r \chi_r d\Omega_{ri} + \sum_j d\left[\lambda_j\Omega_{ji}\right]\right) \text{ for } i \in R+N, \tag{19}$$

where all the variables are indexed by t. If we swap L and χ for L^{comp} and χ^{comp}, we get the evolution of sales shares in the compensated equilibrium instead. The boundary conditions are the same for both the compensated and decentralized equilibria — spending shares are equal to their values in status-quo, and the initial level of prices and wages is irrelevant.

Proposition 2 can be used to calculate either the decentralized equilibrium, given $d\log\chi$ and $d\log L$, or the compensated equilibrium, given $d\log\chi^{\text{comp}}$ and $d\log L^{\text{comp}}$. We now briefly describe the intuition for equations (16)-(19). The expression in (16) is log-linearized input demand by i for input j, which depends on the elasticity of substitution, θ_i, and on how the price of j changes relative to the marginal cost of i. The expression in (17) Shephard's lemma, relating the marginal cost of i to the expenditure-share weighted change in input prices and the productivity shock to i. Equation (18) and (19) are simply log-linearized versions of market clearing conditions (6) and (7).

A simple corollary of Theorem 2 and Proposition 2 is the following.

Corollary 2 (Cobb-Douglas Economies with Common Consumption)**.** *Suppose that all production functions are Cobb-Douglas and that there is a common consumption good in every location. Then,*

$$\log A(t) = \sum_i \lambda_i(0)\Delta\log z_i,$$

regardless of the shape of the $g(c)$ function and distribution of taste and productivity shifters.

If all production functions are Cobb-Douglas and there is a common consumption good, then the right-hand side of (16) and (19) are both zero. That is, changes in location choices and spending shares do not affect the sales shares. In fact, sales shares are constant: $\lambda_i^{\text{comp}}(s) = \lambda_i^{\text{comp}}(0)$. Since compensated variables are equal to decentralized variables in the status-quo, $\lambda_i^{\text{comp}}(0) = \lambda_i(0)$, the integral in Theorem 2 can be solved easily.

4.2 Comparison to Real Output

We contrast changes in aggregate efficiency, defined via $A(t)$, with changes in real output, as measured in the national accounts.

Definition 5 (Real Output)**.** The change in real output is the share-weighted sum of changes in final consumption quantities:

$$\log Y(t) = \int_0^t \int \sum_{i \in N} \frac{p_i(s) c_{hi}(s)}{\sum_{i'} \sum_h p_{i'}(s) c_{hi'}(s)} \frac{d \log c_{hi'}}{ds} dhds,$$

where c_{hi} is household h's consumption of good i.

Using Hulten (1978), we can write real output as a Domar-weighted sum of technology and labor supply changes:

$$\log Y(t) = \int_0^t \left[\sum_i \lambda_i(s) \frac{d \log z_i}{ds} + \sum_r \lambda_r \frac{d \log L_r}{ds} \right] ds.$$

To a first-order approximation, we can write

$$\Delta \log Y \approx \sum_i \lambda_i(0) \Delta \log z_i + \sum_r \frac{w_r}{\sum_{r'} w_{r'} L_{r'}} \Delta L_r.$$

Hence, real output rises if productivities rise, but it also rises if households move from low wage locations to high wage locations.

In contrast, aggregate efficiency, $A(t)$, does not respond to relocations (see Corollary 1). If households choose to move from low to high wage locations, this has no first-order effect on aggregate efficiency. The increase in wages experienced by movers is exactly offset by the reduction in the amenity value of the move. Hence, although real output may rise due to the changes in location, aggregate efficiency does not. We summarize this in the proposition below.

Proposition 3 (First-Order Difference between Aggregate Output and Efficiency)**.** *To a first-order approximation, the difference between aggregate output and aggregate efficiency are changes in income caused by relocation:*

$$\Delta \log Y - \Delta \log A \approx \sum_r \frac{w_r}{\sum_{r'} w_{r'} L_{r'}} \Delta L_r.$$

Hence, real output and aggregate efficiency can differ, even in sign, to a first-order approximation. We illustrate this using the simple one-good economy in Section 3.3.

Example 2 (Output versus Aggregate Efficiency). Consider a positive productivity shock $\Delta \log z_r > 0$ to some region r in the one-good economy in Section 3.3. By Corollary 1, the change in aggregate efficiency is, to a first-order, equal to

$$\Delta \log A \approx \frac{w_r L_r}{\sum_{r'} w_{r'} L_{r'}} \Delta \log z_r > 0,$$

where $w_i L_i$ are evaluated at the initial equilibrium (or final) equilibrium. Hence, an increase in productivity in region r always raises aggregate efficiency by that region's share in sales. However, the change in real output is given by the growth in total consumption:

$$\Delta \log Y = \Delta \log \frac{\sum_{r'} z_r(t) L_r(t)}{\sum_{r'} z_{r'}(0) L_{r'}(0)} \approx \frac{w_r L_r}{\sum_{r'} w_{r'} L_{r'}} \Delta \log z_r + \sum_{r'} \frac{w_{r'} L_{r'}}{\sum_{r''} w_{r''} L_{r''}} \Delta \log L_{r'}.$$

Whereas $\Delta \log A$ always increases in response to a positive productivity shock, $\Delta \log Y$ may decline if wages in r are sufficiently lower than wages in other regions and enough people move to r from other regions in response to the increase in productivity of r. In such a case, total production of the consumption good declines, so $\Delta \log Y < 0$, but every agent is weakly better off.

If we define multifactor productivity as output growth minus the growth in quality-adjusted labor:

$$\Delta \log A^{MFP} = \Delta \log Y - \sum_r \frac{w_r}{\sum_{r'} w_{r'} L_{r'}} \Delta L_r,$$

then $\Delta \log A^{MF} \approx \Delta \log A$ to a first-order. The intuition here is that if we separate changes in output due to technology shocks from changes in output due to relocation, then multifactor productivity, as measured by a national income accountant, coincides to a first-order with aggregate efficiency.

Before concluding this section, we note an important and tractable special case. Suppose there is no amenity value associated with different locations, and suppose there is a single consumption good. In this case, real output and aggregate efficiency coincide.

Proposition 4 (Coincidence of Real Output and Aggregate Efficiency)**.** *Suppose that ϵ_{hr} does not vary by h and r, and that there is a common consumption good. Then*

$$A(t) = A^{MFP}(t) = Y(t).$$

This follows from the fact that under these assumptions, $\boldsymbol{L} = \boldsymbol{L}^{\text{comp}}$, and changes in the regional spending shares, χ^{comp} have no effect on relative prices since all households spend income on the same consumption good (e.g. $d\chi$ drops out of (19)). Hence, $\lambda_i^{\text{comp}}(s) = \lambda_i(s)$. This implies that $A(t) = A^{MFP}(t)$. Furthermore, since every household chooses location to maximize nominal income, households that move from one location to another do not experience a change in their nominal wage, which implies that $d\log Y = d\log A^{MFP}$ — since this holds for any $s > 0$, by integration, it follows that $Y(t) = A^{MFP}(t)$.

The assumptions of Proposition 4 hold in many models of occupational (but not spatial) choice. In such models, a typical assumption is that households vary in their efficiency of different occupations, but have the same preferences across occupations, and consume the same consumption good regardless of their choice of occupation. Under these assumptions, Proposition 4 guarantees that aggregate efficiency, as defined by the aggregate consumption-equivalent variation, coincides with changes in real output, as measured by a national income accountant, in the decentralized equilibrium.

5 Analytical Description of Nonlinearities

Having understood the first-order properties of $A(t)$, we now probe nonlinearities analytically.

5.1 Second-Order Results

Corollary 3 (Second-Order Approximation)**.** *To a second-order approximation, the change in aggregate efficiency satisfies*

$$\Delta \log A \approx \sum_i \lambda_i(0) \Delta \log z_i + \frac{1}{2} \sum_i d\lambda_i^{comp}(0) \Delta \log z_i,$$

where $d\lambda_i^{comp}$ is the change in the compensated sales of i divided by GDP at the status-quo.

The intuition is very similar to the formulas derived by Baqaee and Farhi (2019) for an economy without discrete choice. If sales shares in the compensated equilibrium rise in response to positive productivity shocks, then aggregate efficiency rises more quickly than its first-order approximation would suggest. If sales shares fall, then the reverse happens. Calculating $d\lambda_i^{\text{comp}}(0)$ is a straightforward application of Proposition 2 if we know $dL^{\text{comp}}(0)$ and $d\chi_r^{\text{comp}}(0)$. Hence, in this section, we focus on describing dL^{comp} and $d\chi_r^{\text{comp}}$ at the status-quo, and comparing them to their decentralized uncompensated counterparts.

To simplify the analysis, we focus on the case with only preference heterogeneity and abstract from skill-heterogeneity. That is, in this section, we assume households have homogenous skills: $a_{hr} = a_{h'r'} = 1$. We discuss how to generalize the results to allow for heterogeneity in skills in Appendix C.

Suppose Proposition 5 provides a general characterization of spending shares by region in the decentralized equilibrium, $d\chi(0)$, and the compensated equilibrium $d\chi^{\text{comp}}(0)$.

Proposition 5 (Spending). *In the decentralized equilibrium, the spending share in region r satisfies*

$$d \log \chi_r = d \log w_r - \mathbb{E}_\chi \left[d \log w \right] + d \log L_r - \mathbb{E}_\chi \left[d \log L \right],$$

where the changes in wages $\boldsymbol{w}$ and locations $\boldsymbol{L}$ are in the decentralized equilibrium. In the compensated equilibrium, the share of spending by households in region r at $t = 0$ satisfies

$$d \log \chi_r^{comp} = d \log p_r - \mathbb{E}_\chi \left[d \log p \right] + d \log L_r^{comp} - \mathbb{E}_\chi \left[d \log L^{comp} \right],$$

where the changes in prices $\boldsymbol{p}$ and locations $\boldsymbol{L}^{comp}$ are in the compensated equilibrium.

The intuition is the following. In the decentralized equilibrium, the share of spending by each location must move in line with the share of income earned by households in that location. This is the product of wages w_r and population shares L_r. A region's share of spending rises if wages in that region increase quickly or if households flock to that region. On the other hand, in the compensated equilibrium, the share of spending in each region depends on the rate at which the cost of consumption in region r has increased and the population shares in the compensated equilibrium. If consumption prices rise quickly in a region, or households flock to a region, in the compensated equilibrium, then that region's share of total spending rises.

Proposition 6 provides a general characterization of $dL^{\text{comp}}(0)$ in terms of the derivatives of the uncompensated labor supply function at the status-quo.

Proposition 6 (Migrations). *To a first-order approximation, at $t = 0$, the change in the share of households in region r is*

$$dL_r = \sum_{r' \neq r} dL_{r' \to r} - \sum_{r' \neq r} dL_{r \to r'}, \tag{20}$$

where $dL_{r' \to r}$ is the share of households that move from r to r'. This satisfies:

$$dL_{r \to r'} = - \left[\frac{\partial L_r}{\partial w_{r'} / p_{r'}} d(\frac{w_{r'}}{p_{r'}}) - \frac{\partial L_{r'}}{\partial w_r / p_r} d(\frac{w_r}{p_r}) \right] \mathbf{1} \left[\frac{\partial L_r}{\partial w_{r'} / p_{r'}} d(\frac{w_{r'}}{p_{r'}}) \leq \frac{\partial L_{r'}}{\partial w_r / p_r} d(\frac{w_r}{p_r}) \right], \tag{21}$$

where the changes in wages and prices are the ones in the decentralized equilibrium.

The compensated supply function instead satisfies:

$$dL_r^{comp} = \sum_{r' \neq r} dL_{r' \to r}^{comp} - \sum_{r' \neq r} dL_{r \to r'}^{comp},$$

where $dL_{r' \to r}^{comp}$ *is the share of households that move from* r *to* r' *in the compensated equilibrium. At* $t = 0$*, this satisfies:*

$$dL_{r \to r'}^{comp} = \frac{\partial L_r}{\partial \left[w_{r'} / p_{r'} \right]} \frac{1}{p_{r'}} \left(d(\frac{w_r}{Ap_r})p_r - d(\frac{w_{r'}}{Ap_{r'}})p_{r'} \right) \times \mathbf{1} \left[d(\frac{w_r}{Ap_r})p_r \leq d(\frac{w_{r'}}{Ap_{r'}})p_{r'} \right], \quad (22)$$

where the changes in wages and prices are the ones in the compensated equilibrium and $d \log A$*, at* $t = 0$*, is given by Corollary 1.*

Equation (21) states that in the decentralized equilibrium, households switch from r to r' if the change in L_r due to real wage changes in r' are outpaced by the change in $L_{r'}$ due to real wage changes in r. The overall change in the share of households choosing r, in (20), depends on the sum of all these pairwise switchings. Equation (22) is the counterpart to (21) in the compensated equilibrium. Here, households are given individualized transfers and aggregate factor productivity is adjusted to ensure every household can be made exactly indifferent to the status-quo. In this case, households switch from r to r' if the compensated real wage in dollar terms rises faster in r' than in r, and the rate at which households switch depends on the (uncompensated) cross-elasticity of the share that choose r relative to the real wage in r'.

We illustrate using a simple example below.

Example 3 (Second-Order Approximation of Aggregate Efficiency with Fréchet). Consider again the one good economy described in Section 3.3. We know from Corollary 3 that, to a second-order,

$$\Delta \log A \approx \mathbb{E}_\lambda [d \log z] + \frac{1}{2} \left[Var_\lambda \left(d \log z \right) + Cov_\lambda (d \log L^{\text{comp}}, d \log z) \right], \quad (23)$$

where the variance and covariance operators use λ as the probability weights and we use the fact that $d\lambda_i^{\text{comp}} = d(z_r L_r^{\text{comp}} / (\sum_{r'} z_{r'} L_{r'}^{\text{comp}}))$. In words, the nonlinear change in aggregate efficiency depends on the variance of the shocks and the covariance of compensated migrations with the shocks. The variance term reflects the fact that, even if households are not mobile, an increase in productivity in one location raises that location's sales share, which magnifies the value of positive productivity growth in that region. The covariance reflects the fact that if, in the compensated equilibrium, households move to locations experiencing more rapid productivity growth, then this further boosts the compensated sales shares of producers in that location, and raises aggregate efficiency.

Suppose that household preferences can be written as $u_h(c_h, l) = f_h(\log(c_h) + \epsilon_{hl})$ for some increasing function f_h where ϵ_{hl} are distributed according to a Frechet distribution with tail parameter θ. Under this assumption, the share of households that choose location r given a vector of real wages $\boldsymbol{z}$ is

$$L_r(z) = \frac{B_r z_r^\theta}{\sum_{r'} B_{r'} z_{r'}^\theta}, \tag{24}$$

for some parameters $B_r > 0$. (see, e.g. Redding and Rossi-Hansberg, 2017). Accordingly, the uncompensated cross-derivatives of supply are:

$$\frac{\partial L_{r'}}{\partial w_r} = -\theta L_{r'} L_r \frac{1}{z_r}.$$

For simplicity, suppose that initial productivities, $z_r(0)$ are the same in every location. Then applying Proposition 6 yields

$$d \log L_i^{\text{comp}} = \theta \left(d \log z_i - \mathbb{E}_\lambda \left[d \log z_i \right] \right).$$

In this example, it is easy to verify that changes in population shares in the compensated equilibrium, $d \log L^{\text{comp}}$, are the same as the ones in the decentralized equilibrium, $d \log L$, up to a first-order approximation.

Substituting this into the approximation for $\Delta \log A$ in (23) yields:

$$\Delta \log A \approx \mathbb{E}_\lambda [d \log z] + \frac{1}{2}(1 + \theta) Var_\lambda \left(d \log z \right).$$

Hence, the higher is θ, the more convex is aggregate efficiency.[14]

5.2 Differential Equations

Propositions 5 and 6 allow us to approximate aggregate efficiency $A(t)$ to a second-order approximation given expenditure shares, population shares, and uncompensated elasticities. However, computing changes in aggregate efficiency beyond second-order requires

[14] If instead the share function were logit (i.e. $g(c_h)$ is linear and ϵ_{hr} are type I extreme value with parameter θ^{logit}), then $\Delta \log A \approx \mathbb{E}_\lambda [d \log z] + \frac{1}{2}(1 + \theta^{\text{logit}} z) Var_\lambda \left(d \log z \right)$. Calibrating $\partial L_r / \partial z_{r'}$ to be the same as in the Frechet model implies $\theta^{\text{logit}} = \theta / z$. Hence $\Delta \log A$ would be the same in the two models up to a second-order approximation.

evaluating compensated population L^{comp} and spending shares χ^{comp} away from status-quo at $t > 0$. This can be accomplished via Proposition 1 and simulation methods (i.e. simulate a large population of households and apply Proposition 1). However, there is one noteworthy case where can characterize $\log A(t)$ nonlinearly without using simulation methods. This is when the following assumption holds.

Assumption 1 (Common Consumption Good and Linear g)**.** Suppose that there is one common consumption good in every region, $p_r = p_{r'}$, and denote its price by p_0. Suppose that preferences take the form $u_h(c_h, l) = f_h(c_h + \epsilon_{hl})$ for some arbitrary increasing function f_h.

We show in Appendix A that the requirement that $u_h(c_h, l)$ be representable as $f_h(c_h + \epsilon_{hl})$ implies that the supply function must have symmetric cross-derivatives:

$$\frac{\partial L_r}{\partial(w_{r'}/p_0)} = \frac{\partial L_r'}{\partial(w_r/p_0)}.$$

An example is the logit supply system described in Equation (10). A counterexample is the Frechet supply system in (24).

Proposition 7 (Exact Characterization without Simulation)**.** *If Assumption 1 holds, then the compensated supply function coincides with the (uncompensated) supply function: $L^{comp}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{u}) = L(\boldsymbol{w}, \boldsymbol{p})$. Hence, for any value of $t > 0$, we have*

$$dL_r^{comp} = \sum_{r'} \frac{\partial L_r}{\partial\left[w_{r'}/p_0\right]}\left[d(\frac{w_{r'}}{Ap_0}) - d(\frac{w_r}{Ap_0})\right], \tag{25}$$

where prices and wages are evaluated in the compensated equilibrium at t.

Since there is a common consumption good, the distribution of spending $\chi_r^{\text{comp}}(t)$ has no effect on compensated sales shares $\lambda^{\text{comp}}(t)$. Hence, we can combine Proposition 7 with Theorem 2 and Proposition 2 to compute $A(t)$ as the solution to a system of ordinary differential equations. We provide an example below.

Example 4 (Occupational Choice Model). Suppose consumption is a CES bundle of outputs from different industries:

$$y = \left(\sum_r \Omega_{0r}^{\frac{1}{\theta_0}} x_{0r}^{\frac{\theta_0 - 1}{\theta_0}}\right)^{\frac{\theta_0}{\theta_0 - 1}},$$

where the units of quantities are chosen so that Ω_{0r} are expenditures in the status-quo. Industry r's output is

$$x_{0r} = z_r L_r.$$

Suppose that workers utility functions can be written as

$$u_h(c_h, r) = c_h + \epsilon_{hr},$$

where ϵ_{hr} is type I extreme value. In this case, $L(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{T})$ has the standard logit functional form. This implies that

$$\frac{\partial L_r}{\partial \left[w_{r'}/p_0\right]} = \theta L_r L_{r'}.$$

Substituting this into (25) yields

$$d\log L_i^{\text{comp}} = \theta \left(d\left[w_i/\left(Ap_0\right)\right] - \mathbb{E}_{L^{\text{comp}}}\left[d\left[w_i/\left(Ap_0\right)\right]\right]\right), \tag{26}$$

where the wages and prices are evaluated in the compensated equilibrium. We now apply Proposition 2. The sales share of industry r, in the compensated equilibrium, is given by its share of the wage bill:

$$d\log\lambda_l^{\text{comp}} = d\log w_l + d\log L_l^{\text{comp}} - \mathbb{E}_{\lambda^{\text{comp}}}\left[d\log w + d\log L^{\text{comp}}\right]. \tag{27}$$

At the same time, the sales of share industry r, in the compensated equilibrium, is also given by the share of household spending on industry r:

$$d\log\lambda_l^{\text{comp}} = (\theta_0 - 1)\left[d\log z_l - d\log w_l - \mathbb{E}_{\lambda^{\text{comp}}}\left[d\log z - d\log w\right]\right], \tag{28}$$

where we use the fact that $d\log p_r = d\log w_r - d\log z_r$. Shephard's lemma implies that the consumer price index in the compensated equilibrium is

$$d\log p_0 = \sum_l \lambda_l^{\text{comp}}\left[d\log w_l - d\log z_l\right]. \tag{29}$$

Finally, from Theorem 2, the change in aggregate efficiency is:

$$d\log A = \sum_i \lambda_i^{\text{comp}} d\log z_i. \tag{30}$$

Equations (26)-(30) form a system of ordinary differential equations that can be solved to obtain $A(t)$ without simulation methods. The boundary conditions are that at $t = 0$, expenditure and population shares coincide with the status-quo, and $A(0) = 1$. Solving the system is simple: we discretize the productivity shocks, and iterate on the linear system, updating variables each time.

---

6 Comparison to Other Measures of Aggregate Efficiency

In this final section, we compare our measure of aggregate efficiency with two popular alternatives in the literature. First, we contrast our approach with the popular utilitarian or "expected" utility measure of aggregate welfare. Second, we compare our results to the sum of compensating variations.

6.1 Utilitarian Social Welfare

Definition 6. Define utilitarian social welfare function to be

$$U(\boldsymbol{c}) = \mathbb{E}[\max_{l_h} u_h(c_{l_h}(t), l_h)], \tag{31}$$

where $\mathbb{E}$ is a population-weighted cross-sectional average and $c_{l_h}(t)$ is consumption in location l_h at t. The consumption-equivalent variation for utilitarian welfare is $A^U(t)$ that solves:

$$U(\boldsymbol{c}(t)/A^U(t)) = U(\boldsymbol{c}(0)),$$

The function in (31) is sometimes called "expected utility" but since households do not face lotteries and make choices over taste parameters, this is a misleading label. Instead, as pointed out by Donald et al. (2023), (31) is more appropriately thought of as a particular social welfare function.[15]

Unlike $A(t)$ and $Y(t)$, the average utility measure $A^U(t)$ is not invariant to monotone transformations of individual utility functions. In particular, its value is not pinned down by any observables. To see this, recall that $u_h(c_h, l) = f_h(g(c_h) + \epsilon_{hl})$, where f_h is an arbitrary strictly increasing functions. Changes to f_h have no testable implications in terms of households' choices. In particular, any choice of f_h represents the same underlying preference relation $\succeq_h$. However, altering f_h has important implications for the value of A^U.

[15]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 paper, 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 the ranking of u_h and $u_{h'}$ has no testable or observable implications.

We illustrate this fact using a simple example below.

Example 5 (Average utility not pinned down by observables). Consider a two region example, and suppose agent h has utility function:

$$u_h(c_r, r) = \varepsilon_{hr} c_r,$$

where ε_{hr} are i.i.d Fréchet random variables with tail parameter θ. It is well-known that this implies

$$L_r = \frac{c_r^\theta}{\sum_{r'} c_{r'}^\theta}.$$

As mentioned before, utility functions are only pinned down up to monotone transformations. So, suppose that we instead consider

$$u_h(c_r, r) = \bar{\epsilon}_h \varepsilon_{hr} c_r,$$

where $\bar{\epsilon}_h$ is an h-level shifter. Any conceivable choice data generated by the model is consistent with every value of $\bar{\epsilon}_h > 0$. However, the average utility metric $A^U(t)$ depends on the choice of $\bar{\epsilon}_h$. In particular, using the law of total expectation, we can write

$$A^U(t) = \frac{\mathbb{E}\left[\bar{\epsilon}_h \mathbb{E}\left[\max_{l_h}\{c_{l_h}(t)\varepsilon_{hl_h}\}\middle|\,\bar{\epsilon}_h\right]\right]}{\mathbb{E}\left[\bar{\epsilon}_h \mathbb{E}\left[\max_{l_h}\{c_{l_h}(0)\varepsilon_{hl_h}\}\middle|\,\bar{\epsilon}_h\right]\right]}, \tag{32}$$

If we further assume that the household-level shifter $\bar{\epsilon}_h$ is independent of the taste shifters $\{\varepsilon_{hr}\}$, then, as shown by Redding and Rossi-Hansberg (2017), $A^U(t)$ can be written in closed-form as[16]

$$A^U(t) = \left[\frac{c_1^\theta(t) + c_2^\theta(t)}{c_1^\theta(0) + c_2^\theta(0)}\right]^{\frac{1}{\theta}}. \tag{33}$$

This is a very popular measure of aggregate welfare in spatial models.

However, inspection of Equation (32) shows that the untestable assumption about the joint independence of $\bar{\epsilon}_h$ and ε_{hr} is important. Other assumptions will lead to other social welfare functions that are consistent with the same observables. For example, suppose instead, of assuming $\bar{\epsilon}_h$ is independent of ε_{hr}, we assume that $\bar{\epsilon}_h$ is equal to $1/\mathbb{E}[\varepsilon_{hr}|h]$. That is, we normalize the utility of agent h by the average value of ε_h so that, on average, taste shocks for every household are equal to one (some households cannot have higher taste parameters in every location).

[16]Technically, if $\theta < 1$, then $\mathbb{E}[u_h(c_h(t), l_(t))]$ diverges under the assumption that $\bar{\epsilon}_h$ is independent of ε_{hr}. However, since the ratio is well-defined, $A^U(t)$ is still well-defined.

Under this assumption, (32) is instead,

$$A^U(t) = \frac{D(L_1(t))c_1(t) + D(L_2(t))c_2(t)}{D(L_1(0))c_1(0) + D(L_2(0))c_2(0)},$$

where $D(x) = \int_0^x u^{-\frac{1}{\theta}} / (u^{-\frac{1}{\theta}} + (1-u)^{-\frac{1}{\theta}} du$. Hence, with this alternative untestable and equally plausible assumption about the average level of taste shocks by household, we arrive at different results about $A^U(t)$. Since two assumptions are observationally equivalent, there is no conceivable choice data that can distinguish between these two assumptions.

The previous example shows that $U(\boldsymbol{c})$ depends on an untestable assumption about how individual's ordinal preference relations are represented by utility functions. Nevertheless, one might think that setting $f_h(x) = x$ might result in a $U(\boldsymbol{c}(t))$ which nevertheless is similar to $A(t)$. This turns out to not be the case in general, as the following example shows.

Example 6 (Average utility versus aggregate efficiency). Consider again Example 5. Suppose there is a common freely-traded consumption good produced linearly from labor, with productivity z_r in location r. Under the typical assumption that $\bar{\epsilon}_h$ is independent of ϵ_{hr}, it is well-known that to a first-order approximation, changes in $\log A^U$ are a population weighted sum of changes in productivities:

$$\Delta \log A^U(t) \approx \sum_r L_r(0) \Delta \log z_r.$$

This does not coincide, even to a first-order, with changes in aggregate efficiency as measured by A, because $\Delta \log A \approx \sum_r \lambda_r(0) \Delta \log z_r$ (unless all regions have the same initial productivity). It can easily be the case that $\Delta \log A > 0$, so that winners can compensate losers and have resources left over, but have $\Delta \log A^U < 0$.

In general, $A^U(t)$ has embedded within it some distributional concerns — which get stronger as mobility falls — even though this may not be readily apparent. To see this, consider the following example.

Example 7 (Optimal policy for average utility). Consider again the economy in Example 6. Suppose that the government pursues place-based policy to maximize $U(\boldsymbol{c})$. That is, the government considers setting real consumption in each location according to

$$\max_{c_r} U(c_1, c_2) \tag{34}$$

subject to the constraint that total consumption is feasible:

$$\sum_r L_r c_r = \sum_r L_r z_r$$

and that the share of households in each region is

$$L_r = \frac{c_r^{\theta}}{\sum_{r'} c_{r'}^{\theta}}.$$

One can show that under the common assumption that $\bar{\epsilon}_h$ is independent of ε_{hr}, the optimal c_r is a convex combination of production in r and average production:

$$c_r = \frac{\theta}{\theta+1} z_r + \frac{1}{\theta+1}\left(\sum_r L_r c_r\right). \tag{35}$$

Hence the optimal c_r redistributes income from rich to poor regions. Under the assumption that $\bar{\epsilon}_h$ is $1/\mathbb{E}[\varepsilon_{hr}|h]$, the optimal c_r still redistributes income from rich to poor regions, but by a different amount. That is, different untestable but equally plausible assumptions lead to different amounts of "optimal" place-based redistributive policy.

Of course, by the first welfare theorem, the decentralized equilibrium in this model is Pareto efficient. Hence, if we swap the objective of the government to be to maximize aggregate efficiency, A, then the government would not intervene in the equilibrium allocation since it is already Pareto efficient. In contrast, the "optimal" allocation according to (35) is not Pareto efficient (except in knife edge cases). Intuitively, in this allocation, the marginal product of choosing to live in a region, z_r, is not equal to the payment the marginal household receives. Hence the planner solving the problem in (34) sacrifices efficiency to pursue distributional goals.

There is a case where utilitarian welfare, or the social surplus function, can be used to calculate $A(t)$. This happens if Assumption 1 holds.

Proposition 8 (Using Average Utility to Calculate A)**.** *Suppose that Assumption 1 holds. Let $\boldsymbol{w}(\boldsymbol{z}, Z)$ be real wages in a decentralized equilibrium with productivity parameters $\boldsymbol{z}$ and aggregate factor productivity Z. Then, $A(t)$ solves*

$$U(\boldsymbol{w}(0)) = U(\boldsymbol{w}(\boldsymbol{z}(t), A(t))/A(t)), \tag{36}$$

where $U(c) = \mathbb{E}[\max_{l_h} c_{l_h} + \epsilon_{hl_h}]$.

In words, Proposition 8 provides a set of assumptions under which a utilitarian social welfare function can be used to compute $A(t)$. In particular, $A(t)$ is the reduction in labor productivity in every location such that utilitarian welfare, with productivity shocks $\boldsymbol{z}(t)$, is equal to welfare under the status-quo. Notably, $\boldsymbol{w}(\boldsymbol{z}(t), A(t))$ need not be the same as the real wages in the decentralized equilibrium given the same productivity shifters $\boldsymbol{w}(\boldsymbol{z}(t), 1)$. This is because, as we scale aggregate factor productivity to make compensations feasible, agents may switch locations, which would affect real wages.

The example below illustrates Equation (36) using a simple example.

Example 8 (Using average utility to calculate aggregate efficiency with logit). Consider the single good economy example from Section 3.3. There is a single consumption good produced linearly from labor and productivity in each location r is $\boldsymbol{z}_r$. Given a vector of productivity shifters $\boldsymbol{z}$ and an aggregate factor productivity shifter Z, the real wage per efficiency unit of labor is simply $w_r(\boldsymbol{z}, Z) = z_r$. Suppose that $g(c) = c$ and taste shifters are drawn from a Type I extreme value distribution, so that the uncompensated share function is of the logit type. Then (36) is simply equivalent to Equation (15):

$$U(\boldsymbol{z}(0)) = \sum_r \exp(\theta z_r(0) + B_r) = \sum_r \exp(\theta z_r(t)/A(t) + B_r) = U(\boldsymbol{z}(t)/A(t)),$$

where B_r are exogenous shifters. In this example, $A(t) = A^U(t)$, since $\boldsymbol{w}(\boldsymbol{z}(t), A(t)) = \boldsymbol{w}(\boldsymbol{z}(t), 1)$.

The proof of Proposition 8 is instructive. It must be that, in the compensated equilibrium, net transfers are feasible:

$$\int e_h(w^{\text{comp}}(t)/A(t), u_h^0)dh = 0.$$

We can totally differentiate this function with respect to t, use the envelope theorem, and then integrate with respect to t to get that

$$\int_0^t \sum_r L_r^{\text{comp}}[\boldsymbol{w}^{\text{comp}}(s)/A(s), \boldsymbol{u}^0]ds = 0.$$

That is, the area under the compensated supply function must be zero. Under the maintained assumptions, Proposition 7 implies that we can replace the compensated share function with the uncompensated share function, which does not depend on $\boldsymbol{u}_h^0$:

$$\int_0^t \sum_r L_r[\boldsymbol{w}(\boldsymbol{z}(s), A(s))/A(s)]ds = 0,$$

where we use the fact that compensated real wages at s coincide with real wages in the decentralized equilibrium $\boldsymbol{w}(\boldsymbol{z}(s), A(s))$ with productivity shifters $\boldsymbol{z}(s)$ and total factor productivity $A(s)$. The result then follows from the Williams-Daly-Zachary theorem, which states that integrating the uncompensated share function under the maintained assumptions yields the social surplus function defined in (36).

6.2 Kaldor-Hicks Efficiency

In partial equilibrium contexts, where there is an outside good, like money, a common measure of aggregate efficiency is the sum of compensating variations (Small and Rosen, 1981). This measure, also known as the Kaldor-Hicks measure of efficiency, can be defined as follows:

$$S(t) = -\int e_h(\boldsymbol{p}(t), \boldsymbol{w}(t), u_h^0)dh,$$

where e_h is the net expenditure function — the transfer h needs to attain u_h^0, given prices and wages. The scalar $S(t)$ measures the amount of money left, in terms of the numeraire, after winners exactly compensate the losers, holding prices and wages constant at t.

This measure has undesirable properties in general equilibrium, when wages and prices endogeneous respond to redistribution. The following example demonstrates that a pure transfer, which moves the allocations along the Pareto frontier, can nevertheless cause Kaldor-Hicks efficiency to rise.[17] By continuity, this implies that the consumption possibility set can strictly shrink and yet $S(t)$ may rise.

Example 9 (Redistribution raises Kaldor-Hicks efficiency). Consumption in each location is a Cobb-Douglas aggregator of inputs from different locations:

$$\int c_h \mathbf{1}[l_h = r]dh = x_{rr}^{\alpha} \prod_{r' \in R} x_{rr'}^{(1-\alpha)/R},$$

where $\alpha \geq 0$ controls the degree of home bias in consumption. Output in location r is produced one-for-one from labor:

$$\sum_{r'} x_{r'r} = L_r.$$

Workers' utility functions are

$$u_h(c_h, r) = f_h(c_h + \epsilon_{hr}),$$

[17] This is phenomenon also occurs in general equilibrium models with continuous choice and is known as the Boadway (1974) paradox. See Baqaee and Burstein (2025) for more information in a continuous choice setting.

where f_h is any monotone increasing function. Consider an economy with two symmetric regions and suppose that $\epsilon_{h1} = 1$ and ϵ_{h2} are i.i.d uniform random variables in the interval $[1-a, 1+a]$, where a controls the elasticity of location choices to changes in real consumption. This elasticity falls to zero as a rises (since the density of households that are on the margin of switching between locations goes to zero).

The status-quo is a symmetric equilibrium without transfers. Consider a lump-sum transfer from households in region 2 to region 1. That is, agents that chose location 1 in the status-quo, $l_h(0) = 1$, receive a lump-sum transfer of T_1. The transfer is financed by a lump-sum tax on agents that chose location two in the status-quo, $l_h(0) = 2$. Budget balance requires the lump-sum tax T_2 to be $T_2 = L_1(0)/L_2(0)T_1$.

We solve for the equilibrium with redistribution at t, after the redistribution, and calculate the Kaldor-Hicks efficiency measure, $S(t)$. Figure 1 reports $S(t)$ (relative to nominal GDP, which is the numeraire) for different levels of worker mobility (by varying the parameter a). We measure mobility by the mass of agents that start in location 2, for whom $l_h(0) = 2$, and move to location 1 after the redistribution, for whom $l_h(t) = 1$.[18]

Figure 1: Kaldor-Hicks Efficiency as a function of mobility for pure transfer

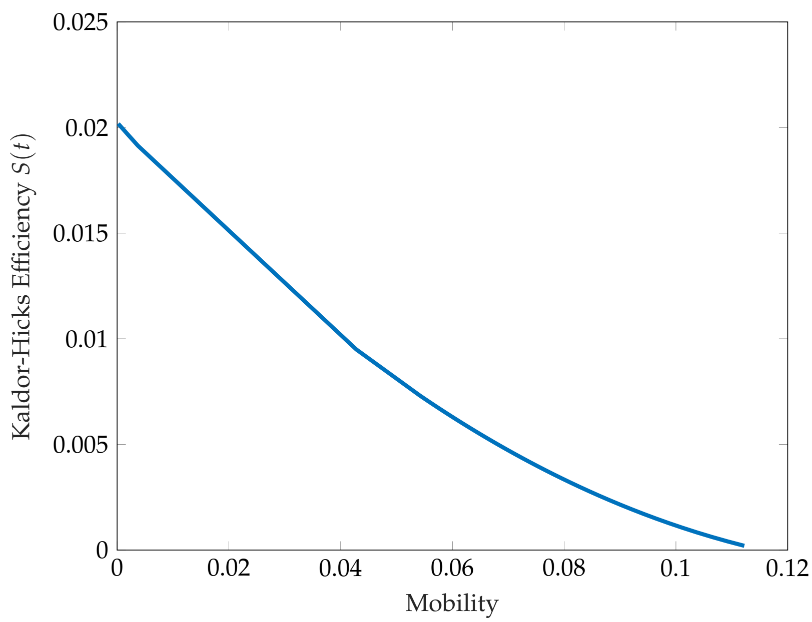

Figure 1 shows that Kaldor-Hicks efficiency, $S(t)$, rises in response to a pure redistribution. In contrast, aggregate efficiency, $A(t)$, measured using the aggregate consumption-equivalent is unchanged at $A(t) = A(0) = 1$ for every value of mobility and preference

[18]In our numerical example, we set the transfer to be 10% of wages in the status-quo $T_1 = 0.1w_1(0)$.

parameters. The reason $S(t)$ is non-zero for pure redistributions is because redistributions change relative wages and prices in general equilibrium. Since $S(t)$ holds wages and prices constant in the compensation, undoing the transfers does not necessarily result in $S(t) = 0$. The magnitude of $S(t)$ depends on parameters — two important parameters are the extent of mobility and home bias. For example, if households do not have heterogeneous tastes, then real wages are equated in the two regions by arbitrage. Since productivities are symmetric, this implies that relative wages and prices are also equated in equilibrium. In this limit, the transfer does not alter relative wages and prices, and so $S(t)$ tends to 0 as we increase mobility.

Similarly, in the absence of home bias ($\alpha = 0$), then $S(t) = 0$. The utility agent h gets in location r is $(w_r + T_h)/p_r$. Without home-bias, the price of consumption in the two locations is the same: $p_1 = p_2$. Hence, because Assumption 1 holds, the transfer T_h does not change h's location choice. That is, $\boldsymbol{L}(t) = \boldsymbol{L}(0)$. Furthermore, since there is no home-bias, agents in both locations spend their income in the same proportion, so that $w_i L_i = 1/2$ (recall that GDP is the numeraire). Since location choices do not respond to the transfers, wages also do not respond to the transfer. Since wages do not respond to the transfers, prices do not respond to the transfers. Since prices and wages do not respond to the redistribution, $S(t) = 0$ in this limit.

Example 9 illustrates that Kaldor-Hicks efficiency is a poor measure of pure efficiency in environments where prices and wages are endogenous to the distribution of income and expenditures.

7 Conclusion

We generalize the cost-benefit approach of Harberger (1971) and Small and Rosen (1981) to measure aggregate efficiency in general equilibrium environments with discrete choice. Our measure converts shocks into a welfare-equivalent change in total factor productivity. The economies studied in this paper are perfectly competitive, Pareto-efficient, and compensations can be made using lump-sum transfers. In ongoing work, we extend the approach in this paper to analyze settings with distortions, externalities, and limited redistribution tools. For an analysis of these issues in general equilibrium economies with heterogeneous agents and continuous choice, see Baqaee and Burstein (2025). Using this approach to analyze optimal policy problems, where maximizing A is the objective function, in distorted settings is another interesting area for future research.

References

Allen, T. and C. Arkolakis (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies 89*(6), 2911–2957.

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

Bagga, S., L. F. Mann, A. Şahin, and G. L. Violante (2025). Job amenity shocks and labor reallocation. Technical report, National Bureau of Economic Research.

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

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

Boadway, R. W. (1974). The welfare foundations of cost-benefit analysis. *The Economic Journal 84*(336), 926–939.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica 87*(3), 741–835.

Davis, M. and J. M. Gregory (2021). Place-based redistribution in location choice models. Technical report, National Bureau of Economic Research.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, 273–292.

Domar, E. D. (1961). On the measurement of technological change. *The Economic Journal 71*(284), 709–729.

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

Dubé, J.-P., J. Joo, and K. Kim (2025). Discrete/continuous choice models and representative consumer theory. *Journal of Economic Theory*, 106012.

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.

Harberger, A. C. (1971). Three basic postulates for applied welfare economics: an interpretive essay. *Journal of Economic literature 9*(3), 785–797.

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

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and us economic growth. *Econometrica 87*(5), 1439–1474.

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.

Lamadon, T., M. Mogstad, and B. Setzler (2022). Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review 112*(1), 169–212.

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

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

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.

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

Redding, S. J. (2025). Evaluating transport improvements in spatial equilibrium. Technical report, National Bureau of Economic Research.

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

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of political Economy 90*(6), 1257–1278.

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.

Appendix A Implications of $g(\cdot)$ for behavior

Since utility is only pinned down up to monotone transformations, the function f_h has no observable implications. However, the shape of $g(c_h)$ has testable implications, since it controls the way income and substitution effects interact with each other. In particular, the share function $\boldsymbol{L}$ uniquely pins down $g(c_h)$ up to an affine transformation.

Proposition 9 (Relation between L and g)**.** *The population share function $\boldsymbol{L}$ pins down the function $g(c_h)$ up to an affine transformation. A notable implication is the following. The population share function without transfers $\boldsymbol{L}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{0})$ has symmetric cross-derivatives in real wages ($\partial L_r / \partial (w_{r'}/p_{r'}) = \partial L_{r'} / \partial (w_r/p_r)$) if, and only if, $g(c_h)$ is a linear function of c_h.*

That is, the functional form of $g(c_h)$ has testable implications. The linear case is noteworthy because, under this assumption, some of our calculations dramatically simplify. Intuitively, if $g(c_h)$ is linear, then a lump-sum transfer (in consumption units) to household h will not change household h's choice of location.[19]

Appendix B Proofs

[To be added.]

Appendix C Extension with Heterogeneous Skills

[To be added.]

[19] As another example, the population share function without transfers $\boldsymbol{L}(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{0})$ has symmetric cross semi-elasticities in real wages $\partial L_r / \partial \log w_{r'}/p_{r'} = \partial L_{r'} / \partial \log w_r / p_r$ if, and only if, $g(c_h)$ is a log function of c_h. The commonly used constant-elasticity share function is a special case and requires that $g(c_h)$ be log.