Aggregate Efficiency with Heterogeneous Agents

David R. Baqaee
UCLA

Ariel Burstein
UCLA*

June 2025

Abstract

In this paper we study aggregate efficiency when households have heterogeneous preferences and outcomes. We generalize the consumption-equivalent variation of Lucas (1987) to a multi-agent setting: we ask by how much can the consumption-possibility set contract while keeping every agent at least indifferent to her status–quo allocation? The resulting scalar equals the resources left over after everyone has been compensated; efficiency rises whenever the same welfare as in the status-quo can be achieved with less resources. We show how every result about computing welfare of a representative agent with homothetic preferences can be converted into a result about aggregate efficiency with heterogeneous (and potentially non-homothetic) preferences. We characterize this measure in terms of initial expenditure shares and elasticities of substitution in both efficient and inefficient economies, and show how to apply our results to study, among other things, the effects of productivity shocks, the gains from trade, and losses from misallocation, with and without costly redistribution.

*We thank Rodrigo Adao, Andrew Atkeson, Natalie Bau, Joao Guerreiro, Valentin Haddad, Oleg Itskhoki, Pablo Fajgelbaum, Yuhei Miyauchi, Gianluca Violante, and Pierre-Olivier Weill for their comments. This document subsumes *Aggregate Welfare and Output with Heterogeneous Agents*, first circulated in 2022.

1 Introduction

A central task of macroeconomics is aggregation: compressing disparate data about the economic activity of firms and households into a handful of numbers, like aggregate output and productivity, that convey useful and welfare-relevant information about the economy. When there is a single final good that can be costlessly transferred across individuals, measuring aggregate output is uncontroversial: simply add up units of that good. Real economies, however, contain many goods consumed by households that differ in tastes, wealth, and the prices they face. This heterogeneity, now central to modern macroeconomics, makes aggregation a hard but essential problem.

The literature that studies heterogeneity in production, e.g. multiple goods, heterogeneous firms, and input-output networks, typically imposes a single, homogeneous final goods aggregator.[1] Under this assumption, changes in aggregate output and efficiency can be unambiguously quantified using the consumption-equivalent variation of Lucas (1987), which asks: by how much must the final consumption bundle be scaled to leave the representative agent indifferent relative to some baseline status-quo. If the answer is less than one, then aggregate efficiency has increased relative to the baseline, because we can make the agent indifferent and still have goods left-over.

By contrast, the literature that focuses on household heterogeneity has not settled on a single measure of aggregate output or efficiency. Although it is common to report aggregate measures, different papers use different metrics. Common approaches to aggregating across households are to: (1) use social welfare functions, (2) sum up compensating variations, or (3) rely on aggregate quantity indices like (chain-weighted) real consumption or GDP. However, each is known to have limitations. For example, social welfare functions hinge on normative choices like Pareto-weights and the degree of inequality-aversion, and so embed subjective (i.e. not falsifiable) interpersonal comparisons. They are also not invariant to monotone transformations of utility functions. On the other hand, the sum of compensating variations, sometimes called Kaldor-Hicks efficiency, suffers from paradoxes in general equilibrium and requires lump-sum transfers to be interpretable. Quantity indices, like chain-weighted real consumption, have pathological properties (like chain-drift) and do not typically account for time and risk.[2]

[1] See, for example, the recent surveys by Hopenhayn (2014), Carvalho and Tahbaz-Salehi (2019), and Baqaee and Rubbo (2023).

[2] We discuss these issues in the paper, but we are not the first to point out these shortcomings. For the paradoxes and limitations of the sum of compensating variations, see Boadway (1974), Blackorby and Donaldson (1990), and Schulz et al. (2023). For pathological properties of chain-weighted quantity indices, see the discussion of chain-drift and path-dependence in Hulten (1973), and Baqaee and Burstein (2023). For how traditional index numbers can be adapted to account for time and risk see, e.g., Alchian and Klein

In this paper, we consider an alternative approach to aggregating across consumers that generalizes Lucas' consumption-equivalent variation, and does not suffer from the limitations described above. To measure changes in aggregate efficiency we ask: by how much can the consumption-possibility set contract while keeping every agent at least indifferent to her status–quo allocation? The resulting scalar measures the resources left over after everyone has been compensated; efficiency rises whenever the same welfare as in the status-quo can be achieved with less resources.[3]

If there is one agent and the consumption-possibility set is a single allocation, then this is exactly the definition used by Lucas (1987). Our measure extends this definition by allowing for multiple agents, with different preferences, and accommodates a wide range of mechanisms through which allocations are decentralized (e.g. competitive markets, search-and-matching, bargaining, imperfect competition, etc.) as well as limits to redistribution (e.g. limited taxes or costly transfers). Our definition also builds on and extends ideas in Allais (1979), Debreu (1951, 1954), and Luenberger (1996).

This measure, which we refer to as the *aggregate consumption-equivalent variation*, has some attractive properties. First, unlike social welfare functions, it introduces no free parameters and answers a counterfactual question in terms of observables. Second, it does not take a normative stance on the optimal distribution of resources across individuals beyond compensating every agent relative to the status-quo. For example, if aggregate efficiency increases and there are extra resources left over after everyone has been compensated, our measure takes no stance on who should get those resources. Third, we show that it does not suffer from the paradoxes that afflict Kaldor-Hicks and chain-weighted indices, nor does it require the availability of lump-sum transfers to be interpretable.

We characterize this measure of aggregate efficiency in terms of observables and generalize well-known results in representative-agent economies to heterogeneous-agent settings. This paper also has two stand-alone companions: Baqaee and Burstein (2025a) and Baqaee and Burstein (2025b), where we apply the framework developed in this paper to study misallocation due to financial market incompleteness (in both closed- and open-economies) and aggregate efficiency in random utility models with discrete choice.

The structure of the paper is as follows. In Section 2, we define changes in aggregate efficiency in abstract terms and present a key result. Theorem 1 shows that the prob-

(1973), Del Canto et al. (2023), Fagereng et al. (2022), and Baqaee et al. (2024).

[3] We refer to this as efficiency since it summarizes the amount of resources that can be saved while attaining indifference. We do not refer to this as a measure of aggregate welfare since aggregate welfare is typically defined via a social welfare function and embeds normative judgement, whereas our measure is simply the answer to a counterfactual question about the amount resource savings over and above what is needed to keep every agent at least indifferent.

lem of calculating aggregate efficiency, using aggregate consumption-equivalents, can be converted into an equivalent fictional utility-maximization problem for a representative agent with homothetic preferences. This forms the basis for all other results in the paper because it allows us to port tools used to study the welfare of representative-agents with homothetic preferences, like Hulten (1978), Harberger (1964), Arkolakis et al. (2012), Petrin and Levinsohn (2012), and Baqaee and Farhi (2019c, 2020) to economies with heterogeneous agents.

The abstract environment in Section 2 does not impose much structure on how consumption possibility sets come about. In Section 3, we specialize the environment to general equilibrium economies with prices, distorting wedges, and lump-sum transfers. In this section, we define some popular alternative measures of efficiency used in the literature, so that we can compare and contrast our approach to these alternatives. Namely, we define real output (using a Divisia or chain-weighted index), Kaldor-Hicks efficiency (which compares total income to the sum of compensating variations), and the welfare of a positive representative agent (if such an agent exists). We establish an important equivalence result: if all households have identical homothetic preferences, and face the same relative prices, then our measure of aggregate efficiency (with lump-sum transfers) coincides with the aforementioned alternatives. However, outside of these common but restrictive assumptions, our measure does not coincide with these alternative measures.

A key result in Section 3 is that our measure of aggregate efficiency can be calculated in general equilibrium settings using the notion of a *compensated* equilibrium. The compensated equilibrium is the general equilibrium of an economy with the same technologies and distortions as the real economy, but where there is a fictional representative agent with homothetic preferences. We show that the welfare of this fictional representative agent measures the change in aggregate efficiency. This is an application of Theorem 1 to settings where allocations are decentralized via general equilibrium, and this result makes it straightforward to use tools and methods from representative-agent economies to analyze aggregate efficiency with heterogeneous agents.

Section 4 restricts attention to perfectly competitive economies (assuming away distortions) where the welfare theorems hold. We show that, to a first-order approximation, Hulten (1978) applies to our measure of aggregate efficiency unaltered. That is, to a first-order, our measure of changes in aggregate efficiency coincide with total factor productivity as measured by the Solow residual. We then derive a version of Hulten (1978) that applies to aggregate efficiency nonlinearly and extend the global characterization in Baqaee and Farhi (2019c) to economies with heterogeneous agents and potentially non-homothetic preferences. We show that changes in aggregate efficiency depend only on expenditure shares and elasticities of substitution in the status-quo. We also generalize the sufficient-statistics of Arkolakis et al. (2012), developed for single-agent economies, to quantify the gains from trade in economies with heterogeneous agents.

In Section 5, we open the door to distortions (but keep lump-sum transfers). We derive versions of Hsieh and Klenow (2009), Petrin and Levinsohn (2012), Harberger (1954, 1964), and Baqaee and Farhi (2020) that apply to distorted economies with heterogeneous agents. In particular, we derive a version of the famous Harberger triangles formula that can be used to quantify misallocation in general equilibrium with heterogeneous agents and with unrestricted preferences. We show that there is a sense in which misallocation losses in the heterogeneous agent model are lower than in a representative agent model. We show that our Harberger triangles formula naturally discards dispersion in wedges that occur across households, since such wedges are pure transfers and do not represent (Pareto) inefficiency.

In Section 6, we consider economies without lump-sum transfers. We discuss how Theorem 1 can be used to analyze changes in aggregate efficiency when redistributive instruments are limited. We show that, starting in perfect competition, the change in aggregate efficiency due to a change in primitives is, to a first-order, the same as Hulten (1978). To a second-order, the change in efficiency is equal to what would have happened with lump-sum transfers (characterized in Sections 4 and 5) minus the additional Harberger triangles caused by inefficient redistribution (which are zero if lump-sum transfers are available). We provide some worked-out examples to show how costly redistribution alters the gains from trade relative to autarky and the benefits from skill-biased technical change.

Throughout the paper, we use prices and quantities in the compensated equilibrium to characterize aggregate efficiency in different ways. Since the compensated equilibrium is just the general equilibrium of a fictional representative agent economy, a relatively well-understood object, we defer a systematic characterization of the compensated equilibrium to the end of the paper. Section 7 provides an explicit characterization of how prices and quantities change in the compensated equilibrium as a function of elasticities of substitution, initial expenditure shares, initial wedges, and changes in primitives.

Related literature. Our approach to measuring aggregate efficiency is related to willingness-to-pay based measures and have a very long history dating all the way back to at least Dupuit (1844). For example, the compensating variation, and sum of compensating variations, in Hicks (1939) and Kaldor (1939), are special cases. Furthermore, the notion of social surplus in Allais (1979), the coefficient of resource utilization in Debreu (1951, 1954), the measure of efficiency in Farrell (1957), and the benefit function in Luenberger (1996) are all related to our measure. Our contribution relative to these works is to provide a characterization without assuming either Pareto efficiency or even markets (Theorem 1), to explicitly allow for limited or costly redistribution, and to apply our measure to modern models.

Our paper is also related to cost-benefit analysis, typically performed by using the sum of compensating variations, as in Harberger (1971), and related ideas like the marginal value of public funds, Hendren and Sprung-Keyser (2020). The idea behind these measures is to ask: "after the winners compensate the losers using lump-sum transfers, is there still money left on the table?" Our measure of efficiency coincides with these measures when both welfare theorems hold and the consumption-possibility set is linear. However, outside of these cases, the two measures are different. First, if the consumption-possibility set is nonlinear, then as shown by Boadway (1974), a pure transfer between agents causes the sum of compensating variations to exceed initial aggregate income. Intuitively, the transfer lowers prices for goods that are relatively more valued by losers than winners. Hence, it is possible to compensate the losers using the post-transfer prices and still have money left-over. Our measure, which can be defined even when prices do not exist, does not have this property.

Second, unlike the sum of compensating variations, our measure does not presuppose that lump-sum transfers are feasible. In this sense, our approach has similarities to Schulz et al. (2023), who generalize the sum of compensating variations to allow for limited redistribution.[4] Our paper complements and differs from Schulz et al. (2023) in many ways, the most important being a difference in focus. They consider economies with a single consumption good, focusing their attention on a mechanism design problem where lump-sum taxes are unavailable because of asymmetric information. Although our formalism and definitions can be applied to such economies, we do not focus on these issues. Instead, we focus on allowing for multiple goods and heterogeneity in preferences and relative prices faced by consumers. This means that even with perfect information and lump-sum transfers, there are interesting questions about how to aggregate across consumers that consume and value different goods. Even when redistribution is restricted, we do not explicitly consider mechanism design problems.

A different approach to aggregation altogether, following Bergson (1938) and Samuelson (1983), is to use a social welfare function to evaluate outcomes. A prominent example

[4]In response to a shock, they consider a tax reform that makes households indifferent to the status-quo and then measures the monetary value of aggregate welfare gains or losses by the fiscal surplus from this reform.

is the behind the veil-of-ignorance measure of Harsanyi (1955). Social welfare functions are by far the most common approach in the modern literature to aggregating across heterogeneous agents.[5] Our paper, which instead looks for and quantifies the potential for Pareto improvements (i.e. compensating everyone and looking to see if resources are left over), provides an alternative to this methodology.

Following in the social-welfare-function tradition, a recent set of papers, including Bhandari et al. (2021), Dávila and Schaab (2022, 2023), and Donald et al. (2023) provide first-order decompositions of changes in aggregate welfare using social welfare functions comparing. Our goal in this paper is different: we do not provide decompositions of social welfare functions, but instead, define and characterize aggregate efficiency directly as an answer to a counterfactual question. The decompositions in the papers mentioned above contain components the authors refer to as capturing efficiency. However, since our objective is different, our notion of efficiency is also generically different to the efficiency components in these papers. Defining efficiency directly, instead of as part of an approximate decomposition, is useful because it means that we can study large changes.[6]

In terms of the tools and methods, our paper is closely related to the literature that studies the macroeconomic consequences of microeconomic productivity changes and wedges. For productivity changes, this includes Gabaix (2011), Acemoglu et al. (2012), Baqaee and Farhi (2019c) and others. For wedges, this includes Harberger (1954), and more recently, Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Bigio and La'O (2016), Liu (2017), Baqaee and Farhi (2020), among others. We relax the assumption typically maintained in both of these literatures that households have common preferences and face common prices.[7]

Finally, our paper is also related to the gains from trade with heterogeneous agents, since one of our running examples considers the costs of autarky. Much of the work on in-

[5]There is a branch of the literature that assumes observed allocations can be rationalized by maximizing some social welfare function within some parametric class, estimates this function, and uses it to conduct policy analysis (see Heathcote and Tsujiyama, 2021 and the references therein). This is equivalent to assuming there exists a normative representative agent: a hypothetical single decision-maker whose utility function is maximized by observed allocations (Chapter 4 Mas-Colell et al., 1995). Our approach is different since we do not need to assume the existence of either a positive nor normative representative agent. Furthermore, even if a normative representative agent exists, there is nothing to say that its preferences are privileged over any other social welfare function (see Example 6 below).

[6]Whereas infinitesimal changes in our measure of efficiency can be integrated to study large changes, integrals of components of social welfare are path-dependent. To see this point, suppose we approximately decompose changes in some function $y = f(x_1, x_2)$ into $dy \approx (\partial f/\partial x_1)dx_1 + (\partial f/\partial x_2)dx_2$. Then we can write $\Delta y = \int(\partial f/\partial x_1)dx_1 + \int(\partial f/\partial x_1)dx_1$ but, unless $f(x_1, x_2)$ is linear in x_1 and x_2, the size of each component of this nonlinear decomposition depends on the arbitrary path of integration.

[7]One exception is Bornstein and Peter (2024), who study misallocation with differences in tastes and markups across households. In their setting, symmetry and the law-of-large numbers, implies that every households' problem is identical despite the fact that households have different preferences.

ternational trade with heterogeneous agents focuses on the distributional effects of trade. Some examples of papers that also calculate aggregate welfare are Antras et al. (2017) and Galle et al. (2023) (using an Atkinson (1970)-style social welfare function with inequality aversion), Kim and Vogel (2020) (using first-order changes in the sum of compensating variations), and Rodríguez-Clare et al. (2022) (using a population-weighted average of welfare gains across regions), all of which differ from our measure of aggregate efficiency.

2 Abstract Definition and Characterization

We consider an economy populated by households indexed by $h \in \{1, ..., H\}$. Household h has ordinal preferences $\succeq_h$ over commodity vectors $\boldsymbol{c}_h \in \mathbb{R}^N$, where N is the number of goods.[8] Assume preferences are represented by utility functions $u_h(\boldsymbol{c}_h)$,[9] indirect utility functions $v_h(\boldsymbol{p}, I_h)$ and expenditure functions $e_h(\boldsymbol{p}, u)$, where $\boldsymbol{p}$ and I_h denote prices and income/wealth. Let $\boldsymbol{c} \in \mathbb{R}^{H \times N}$ denote a matrix whose hth row, denoted by $\boldsymbol{c}_h$ equals the consumption vector of household h. We refer to $\boldsymbol{c}$ as a consumption allocation.

Fix some consumption allocation, denoted by $\boldsymbol{c}^0$, as the *status-quo*. Consider some feasible consumption allocation set, $\mathcal{C} \subset \mathbb{R}^{H \times N}$. The feasible set $\mathcal{C}$ is a function of deeper primitives, like technologies and policies. We wish to define a scalar that measures the change in efficiency, relative to the status-quo, from using technologies or having policies that give rise to the feasible allocation set $\mathcal{C}$.

There are two challenges that must be overcome. First, since preferences are ordinal and utility functions are only defined up to monotone transformation, our measure of aggregate efficiency cannot be measured in terms of utils (which are meaningless). This requires picking a "unit" of account for aggregate efficiency. Second, different feasible allocations in $\mathcal{C}$ are viewed differently by different households — each allocation entails taking a stance on how surplus or loss is divided among households. To avoid both of these problems, we build on an approach to measuring aggregate efficiency pioneered by Allais (1979), Debreu (1951), Debreu (1954), Lucas (1987), and Luenberger (1996).

Definition 1 (Aggregate Consumption Equivalent Variation)**.** We measure aggregate efficiency by the *Aggregate Consumption-Equivalent Variation* — the largest uniform contraction (or smallest uniform expansion) of the feasible consumption set such that it is possible to make every household indifferent between the contracted allocation and the status-

[8] We assume that preferences are continuous and locally nonsatiated.

[9] That is, for each household $u_h(\boldsymbol{c}_h) \geq u_h(\boldsymbol{c}'_h)$ if, and only if, $\boldsymbol{c}_h \succeq \boldsymbol{c}'_h$. When preferences are homothetic (i.e. whenever $\boldsymbol{c}_h \sim_h \boldsymbol{c}'_h$ then $\lambda \boldsymbol{c}_h \sim \lambda \boldsymbol{c}'_h$ for every $\lambda > 0$), then we can represent preferences with a utility function that satisfies $u_h(\lambda \boldsymbol{c}_h) = \lambda u_h(\boldsymbol{c}_h)$.

quo allocation. Formally,

$$A(\boldsymbol{c}^0, \mathcal{C}) \equiv \max \left\{ \phi \in \mathbb{R} : \text{there is } \boldsymbol{c} \in \phi^{-1}\mathcal{C} \text{ and } u_h(\boldsymbol{c}_h) \geq u_h(\boldsymbol{c}_h^0) \text{ for every } h \right\}. \quad (1)$$

We refer to A as aggregate efficiency throughout the paper. The change in aggregate efficiency relative to the status-quo is

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

For concreteness, say, $\Delta \log A = 0.2$, then this means that after making everyone at least as well off as in $\boldsymbol{c}^0$ there are 20 log points (or $\exp(0.2) - 1 \approx 22\%$) of every good left over to be distributed as desired. Agents may not be consuming the same bundle as in the status-quo after they are compensated — we only require that they be indifferent to the status-quo. If there is a single household and $\mathcal{C}$ is a consumption allocation, then A is the same as the consumption-equivalent notion define by Lucas (1987). As in both Lucas (1987) and Debreu (1951), $\Delta \log A$ treats all commodities symmetrically by shifting the consumption possibility set, $\mathcal{C}$, proportionately in every dimension.[10]

Figure 1 illustrates the change in aggregate efficiency for a simple economy with two households, indexed by h and h', and two consumption goods, one consumed only by h and the other only by h'. The set $\mathcal{C}$ could be, but need not be, the technological Pareto frontier. As the figure shows, if $\Delta \log A > 0$, then it must be the case that there are feasible allocations in $\mathcal{C}$ that Pareto-dominate the status-quo $\boldsymbol{c}^0$.

The measure in Definition 1 has some desirable properties: (1) it answers a counterfactual question about observable phenomena with interpretable units. This means it is invariant to monotone transformations of utility functions, and only relies on ordinal properties of preference relations. (2) Our measure does not take a stance on how social

[10]Note that such a choice is present even with a single agent. For example, do we shrink or expand every element in a consumption vector to reach indifference, as in the consumption-equivalent variation in Lucas (1987), or only some components. Similarly, do we scale the whole budget set proportionally, as in the compensating variation of Hicks (1939), or do we change individual prices of certain commodities to reach indifference. In this sense, this choice is not special to having heterogeneous agents or not. In general equilibrium, expanding and contracting the feasible set proportionally is isomorphic to scaling the vector of factor endowments (including quasi-fixed factors capturing decreasing returns to scale). This is an intuitive way to measure efficiency — reductions in necessary factor endowment quantities in order to reach the same status-quo. Our measure of aggregate efficiency can differ from standard approaches like using social welfare functions or chained real consumption even if there is only a single consumption good (when there is no choice of how to expand or shrink the consumption set). To see an example, see Baqaee and Burstein (2025a) where there is a single consumption good in a model with mobility and amenity. Lastly, in some settings, it may be interesting to expand or shrink $\mathcal{C}$ in a particular direction, rather than radially. This requires generalizing our definition to allow for non-radial expansions, as in the transferable surplus notion in Allais (1979). We do not pursue this generalization in this paper.

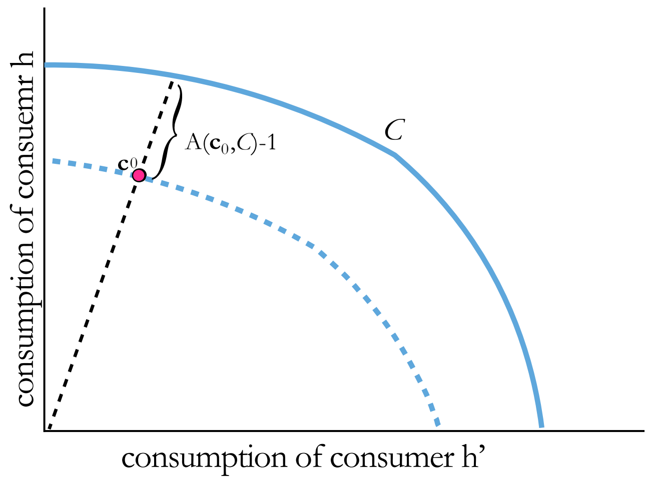


Figure 1: Aggregate efficiency is measured by the maximal radial expansion of the feasible set necessary to achieve indifference.

surplus or losses should be divided among agents. That is, while we can assign a numerical efficiency value to every feasible set of consumption allocations, $\mathcal{C}$, we do not attempt to pick a specific allocation among the possibilities as being socially "optimal." (3) Our definition places no restrictions on the set of redistributive tools that are available or the mechanism by which allocations are decentralized (e.g. spot markets, search, bargaining, etc.). The instruments of redistribution are implicitly embedded into the size of the feasible set $\mathcal{C}$.[11] (4) This abstract definition is applicable to a wide variety of models and circumstances, allowing for, among other things, non-homothetic preferences, discrete choice, risk, and dynamics.

It should be noted that $\Delta \log A$ is not a social welfare function. A social welfare function is an increasing function of individual utilities. Our measure of aggregate efficiency does not satisfy this requirement. First, it is not an increasing function of the underlying utilities, and in fact, only depends on the initial indifference curve of each household (not the entire preference relation). Second, $\Delta \log A$ depends on the status-quo, which means that it is not stable across comparisons if the status-quo changes. Third, $\Delta \log A$ does not take a stance on which point in $\mathcal{C}$ is "socially" optimal, unlike a social welfare function which typically selects some specific allocation in $\mathcal{C}$ as the optimal one.

Aggregate efficiency gains are higher for larger sets: If $\mathcal{C}' \subseteq \mathcal{C}$, then $\Delta \log A(\boldsymbol{c}^0, \mathcal{C}') \leq \Delta \log A(\boldsymbol{c}^0, \mathcal{C})$. For example, if $\mathcal{C}'$ is a convex possibility set from a neoclassical economy and $\mathcal{C}$ is the aggregate budget set for equilibrium prices of the same economy (so that $\mathcal{C}$

[11]For example, for convex competitive economies, if lump-sum transfers are available, then the second welfare theorem implies that $\mathcal{C}$ is the set of all Pareto-efficient allocations. However, the definition can also be applied to economies where such transfers are not available.

is a hyperplane that is tangent to $\mathcal{C}'$) then, given some status-quo, aggregate efficiency gains are lower (or losses are higher) under $\mathcal{C}'$ than under $\mathcal{C}$. A similar logic applies when the curvature of $\mathcal{C}$ is shaped by factors that cannot be reallocated across producers (e.g. sector-specific factors). The more curved $\mathcal{C}$ is, the smaller the efficiency gains are.[12]

To characterize $\Delta \log A$, we prove a very useful theorem, which we make repeated use of throughout the rest of the analysis. This theorem proves that calculating $\Delta \log A$ is equivalent to solving a utility-maximization problem for some fictitious representative agent. This result allows us to translate and port representative-agent results to environments with heterogenous agents.

To do this, we begin by defining *homothetized* transformations of individual preferences.[13]

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

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

The homothetized utility function, $\tilde{u}_h$, is homogenous of degree one in consumption by construction. If the preference relation $\succeq_h$ is homothetic, then $\tilde{u}$ is a cardinalization of $\succeq_h$ — in this case, $\tilde{u}_h$ ranks consumption bundles in the same order as $\succeq_h$. By construction, $\tilde{u}_h$ is homogenous of degree one and normalized to equal to 1 at $\boldsymbol{c}_h^0$. Consider the following simple example.[14]

Example 1 (Single good). Suppose there is a single consumption good, so $u_h(c_h)$ is a strictly increasing function of the scalar c_h. In this case,

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

regardless of the functional form of u_h. That is, $\tilde{u}_h$ is simply the proportional change in the consumption good relative to the status-quo.

---

[12]Our measure of aggregate efficiency can be used to order feasible sets under a given status-quo. The ordering of two feasible sets may flip for different status-quos, similar to Scitovszky (1941). For a fixed status-quo, our measure of aggregate efficiency gives a unique ordering of feasible sets.

[13]The homothetized utility function is also called the *distance* function in the duality literature on optimization (see, for example, Cornes, 1992).

[14]The magnitude of $\tilde{u}_h(\boldsymbol{c}_h)$ is interpretable — it measures the amount the consumption bundle $\boldsymbol{c}_h$ has to be scaled to make the household exactly indifferent to the status-quo. In this sense, $\tilde{u}_h(\boldsymbol{c}_h)$ is the household's consumption equivalent variation relative to the status-quo.

If $\succeq_h$ is non-homothetic, then $\tilde{u}_h$ is *not* a cardinalization of $\succeq_h$ (i.e. $\tilde{u}_h$ does not rank consumption allocations according to $\succeq_h$). Figure 2 graphically depicts indifference curves of $\tilde{u}_h$ — they are radial expansions of the status-quo indifference curve defined by $u_h(\boldsymbol{c}_h) = u_h(\boldsymbol{c}_h^0)$. When $\succeq_h$ is homothetic, all indifference curves are radial expansions, so that the ranking produced by $\tilde{u}_h$ coincides with the one produced by $\succeq_h$. We provide a non-homothetic example below.

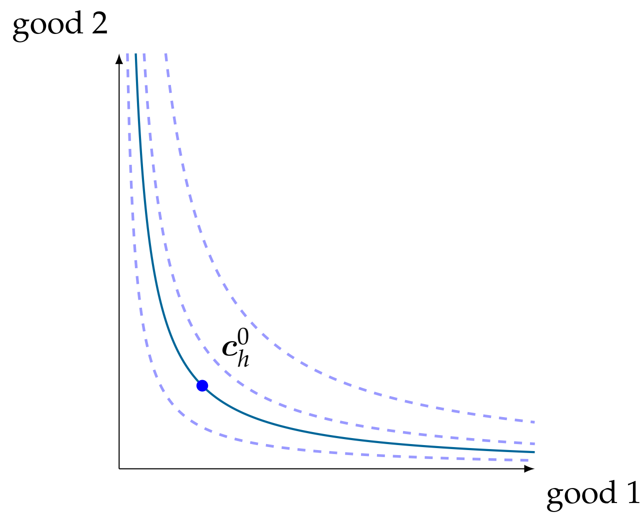


Figure 2: The solid blue line is the indifference curve $u_h(\boldsymbol{c}_h) = u_h(\boldsymbol{c}_h^0)$ and the dashed lines are the indifference curves of $\tilde{u}_h$.

Example 2 (Non-homothetic CES). Consider a household with non-homothetic CES preferences, as in Comin et al. (2021),

$$u_h(\boldsymbol{c}_h) = \left(\sum_i (c_{hi})^{\frac{\eta-1}{\eta}} (u_h(\boldsymbol{c}_h))^{\xi_i} \right)^{\frac{\eta}{\eta-1}}.$$

where η is the compensated elasticity of substitution and ξ_i controls income effects. Then $\tilde{u}_h(\boldsymbol{c}_h)$ is homothetic CES given by

$$\tilde{u}_h(\boldsymbol{c}_h) = \frac{1}{u_h^0} \left(\sum_i (c_{hi})^{\frac{\eta-1}{\eta}} \left(u_h^0\right)^{\xi_i} \right)^{\frac{\eta}{\eta-1}},$$

where $u_h^0 \equiv u_h(\boldsymbol{c}_h^0)$ is a treated as a constant. If ξ_i are the same for every i, then $\tilde{u}_h$ and u_h are both cardinalizations of the same preference rankings.

We now the concept of a fictitious Hicksian representative agent.

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

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

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

The Hicksian representative agent has homothetic preferences and $U(\boldsymbol{c})$ is homogeneous of degree one by construction. The Hicksian representative agent is a fictitious but useful theoretical construct. We call this agent the Hicksian representative agent because if there is only one household, then the demand curves generated by $U(\boldsymbol{c})$ are the Hicksian or compensated demand curves for the single household at the status-quo.[15]

Note that, as was the case with $\Delta \log A$, the function $U(\boldsymbol{c})$ is also not a social welfare function. There are two reasons for this. First, U does not depend on households' true utility functions, instead it depends on the "homothetized" utility functions. Second, U depends on the minimum growth in homothetized utility relative to status-quo, rather than the level of utility. Therefore, despite appearing similar, U is different to the Rawlsian social welfare function, which is the minimum utility levels across all agents.

The following is a main result, because it is used to established almost every other result in the paper.

Theorem 1 (Aggregate Efficiency by Utility Maximization). *Aggregate efficiency is equal to the value of $\mathcal{C}$ to the Hicksian representative agent:*

$$A(\mathcal{C}, \boldsymbol{c}^0) = \max_{\boldsymbol{c} \in \mathcal{C}} U(\boldsymbol{c}).$$

Figure 3 graphically illustrates the content of Theorem 1. Rather than proportionally shifting $\mathcal{C}$ to reach the indifference point, Theorem 1 states that we can instead maximize $U(\boldsymbol{c})$ — by shifting out the indifference curves of the Hicksian representative agent — until reaching the boundary of $\mathcal{C}$. The utility of this fictional agent is numerically identical to the maximal reduction in $\mathcal{C}$ needed to reach indifference.

Theorem 1 is crucial because it converts the problem of calculating aggregate efficiency in (1) into an equivalent utility-maximization problem. Since utility-maximization problems are common in economics, this means that Theorem 1 allows us to easily convert results about representative agent problems into ones about aggregate efficiency with heterogeneous agents.

[15]Given the interpretation of $\tilde{u}(\boldsymbol{c_h})$ in Footnote 14, $U(\boldsymbol{c})$ is the minimum (across households) of each household's consumption equivalent variation.

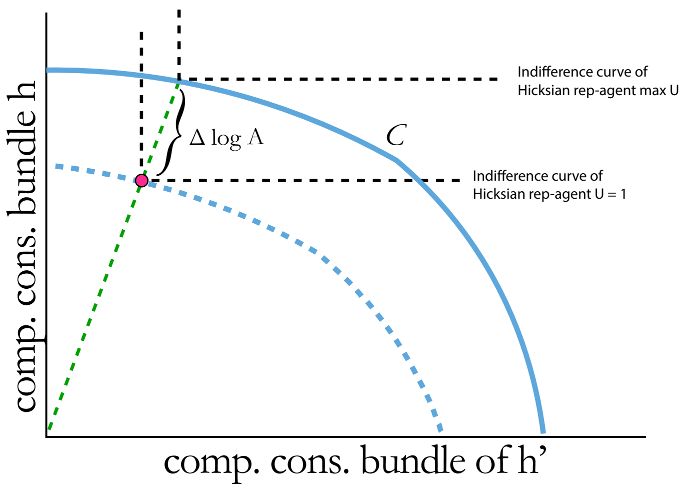


Figure 3: The increase in the utility of the Hicksian representative agent also measures the amount by which $\mathcal{C}$ needs to shrink to ensure indifference.

Theorem 1 guarantees that maximizing $U(\boldsymbol{c})$ yields the same number as the maximization problem that defines A. However, unlike a social welfare maximization problem, the specific allocation in $\mathcal{C}$ that maximizes $U(\boldsymbol{c})$ has no special significance, since the primitive problem defining $\Delta \log A$ is stated in terms of shrinking the possibility frontier, not choosing an allocation inside it.[16].

The rest of the paper uses Theorem 1 in different contexts to characterize changes in aggregate efficiency in terms of observables. We begin by considering productivity changes in competitive economies, before turning attention to inefficient and non-neoclassical economies.

3 General Equilibrium with Distortions and Transfers

In this section we set up a general-equilibrium framework that admits both distortive wedges and lump-sum transfers. We specialize our aggregate consumption-equivalent variation measure of efficiency to this environment and show, via Theorem 1, how to compute it. For comparison we introduce three alternative metrics that are popular in the literature—chain-weighted real output, the sum of compensating variations (Kaldor-

[16]In the special case where $\mathcal{C}$ is a singleton allocation (i.e. no redistribution is possible), denoted by $\boldsymbol{c}'$, then Theorem 1 implies that $\Delta \log A$ is the minimum of the consumption-equivalent change in welfare across all agents comparing the status-quo, $\boldsymbol{c}^0$, to the alternative, $\boldsymbol{c}'$. A less extreme case is when redistribution is possible among only some subset of households. For example, suppose that country h is in autarky from the rest of the world. In this case, a Pareto improvement among all countries except h, leaving h on the same indifference curve, yields $\Delta \log A = 0$, since there is no way to shrink the consumption possibility set radially without making h worse off. In such a case, we may want to redefine $\Delta \log A$ so that the shift in $\mathcal{C}$ excludes goods for the household in autarky (i.e. shrinking and expanding $\mathcal{C}$ only along some dimensions).

Hicks/Cost-Benefit), and the welfare of a positive representative agent—and establish, under restrictive conditions, when all four measures coincide. The remainder of the paper examines the more general cases in which those conditions fail, so our measure no longer aligns with the conventional ones.

3.1 Environment and Equilibrium

Let $u_h(\boldsymbol{c}_h)$ represent household h's preferences. Each household maximizes utility subject to the budget constraint

$$\sum_i \mu_{hi} p_i c_{hi} \leq \sum_f \omega_{hf} w_f L_f + T_h,$$

where the left-hand side is total expenditures and the right-hand side is total income. As in Arrow-Debreu, commodities could be indexed by time and state of nature. On the left-hand side, μ_{hi} is the gross tax rate household h faces on good i, p_i is the price of i not including the tax h pays, and c_{hi} is the quantity of good i purchased by household h. On the right-hand side, households derive income from factors and lump-sum transfers. Households h owns a share ω_{hf} of factor f, where w_f is the wage and L_f is the total quantity of factor f. Lump-sum transfers are T_h.

Producer i chooses its inputs to minimizes costs

$$\sum_j p_j y_{ij} + \sum_f w_f l_{if},$$

subject to production technology

$$y_i = z_i F_i \left(\{y_{ij}\}, \{l_{if}\} \right),$$

where y_i is the quantity of output, F_i is a constant-returns production function, y_{ij} are intermediate inputs used by i produced by j, and l_{if} are primary factors used by i. The assumption that F_i has constant-returns is without loss of generality, since we can capture decreasing returns using producer-specific factors. The parameter z_i is a Hicks neutral productivity shifter. The price of i is equal to an exogenous markup or tax, $\mu_i > 0$, times i's marginal cost of production. That is, the price of i is inclusive of the wedge on i's output.

The assumptions that z_i is Hicks neutral and that the only wedges are on gross output are made without loss of generality. We can capture non-neutral productivity changes, say on i's use of an input from j, by introducing a fictitious intermediary, between i and j, and changing its productivity. Similarly, to capture an implicit tax or wedge on i's use of input j, we can place a markup on that same intermediary. We make the assumption that z_i is Hicks neutral and assume all wedges take the form of taxes on gross output to simplify the notation.

The resource constraint for goods and factors is

$$\sum_j y_{ji} + \sum_h c_{hi} \leq y_i, \qquad \text{and} \qquad \sum_i l_{if} \leq z_f L_f,$$

where z_f, when f indexes a factor, controls the endowment of efficiency units of factor f. Finally, net transfers across households are equal to the revenues generated by the wedges:

$$\sum_h T_h = \sum_i p_i y_i \left(1 - \frac{1}{\mu_i}\right) + \sum_{h,i} (\mu_{hi} - 1) p_i c_{hi}. \tag{2}$$

We define a general equilibrium with wedges below.[17]

Definition 4 (General Equilibrium with Wedges)**.** A *general equilibrium with wedges* is the collection of prices and quantities such that: (1) the price of each good i equals its marginal cost times a wedge μ_i; (2) each producer chooses quantities to minimize costs taking prices as given; (3) each household chooses consumption quantities to maximize utility taking prices, consumption taxes, and income as given; (4) net transfers across households are equal to wedge revenues; (5) all resource constraints are satisfied.

3.2 Different Aggregate Measures in General Equilibrium

We compare our measure of aggregate efficiency, defined in Equation (1), to some popular measures commonly used in the literature. To do so, index exogenous productivity parameters, $\boldsymbol{z}(t)$, wedges, $\boldsymbol{\mu}(t)$, and lump-sum transfers, $\boldsymbol{T}(t)$ by some scalar t and let $t = 0$ denote the status-quo allocation.[18] For any equilibrium price or quantity X, we write $X(t)$ to denote its dependence on the exogenous parameters, $\boldsymbol{z}(t)$, $\boldsymbol{\mu}(t)$, and $\boldsymbol{T}(t)$.[19]

We begin by defining our measure of efficiency.

Aggregate Consumption-Equivalent Variation. Denote the equilibrium consumption allocations given productivity parameters, $\boldsymbol{z}$, wedges, $\boldsymbol{\mu}$, and lump-sum transfers, $\boldsymbol{T}$ by

[17] This notion of general equilibrium is the same one used by Baqaee and Farhi (2020), extended to allow for multiple households.

[18] We abstract from changes in factor ownership shares, but it would be straightforward to allow for them to change exogenously.

[19] In the case of multiple equilibria, we assume there is an equilibrium selection mechanism. The nature of this equilibrium selection mechanism is not relevant for $A(t)$, because $A(t)$ is unique given t and the status-quo.

$c(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{T})$. If the equilibrium is unique, then this is a singleton. Consider the consumption possibility set

$$\mathcal{C}(\boldsymbol{z}(t), \boldsymbol{\mu}(t)) = \{\boldsymbol{c}(\boldsymbol{z}(t), \boldsymbol{\mu}(t), \boldsymbol{T}) \text{ for some transfers } \boldsymbol{T} \text{ satisfying (2)}\}.$$

In words, $\mathcal{C}(\boldsymbol{z}(t), \boldsymbol{\mu}(t))$ is the set of equilibrium consumption allocations that can be attained by varying lump-sum transfers. The second welfare theorem states that $\mathcal{C}(\boldsymbol{z}, \boldsymbol{1})$ is the Pareto-frontier — when there are no wedges, any point on the Pareto frontier can be supported as the outcome of general equilibrium with appropriate lump-sum transfers. If there are wedges, $\boldsymbol{\mu} \neq \boldsymbol{1}$, then the distorted consumption possibility set $\mathcal{C}(\boldsymbol{z}, \boldsymbol{\mu})$ is strictly contained in the undistorted consumption possibility set $\mathcal{C}(\boldsymbol{z}, \boldsymbol{1})$.

Applying Definition 1 with respect to $\mathcal{C}(\boldsymbol{z}(t), \boldsymbol{\mu}(t))$, we obtain our measure of aggregate efficiency:

$$A(t) = \max\left\{\phi \in \mathbb{R} : \text{there is } \boldsymbol{c} \in \phi^{-1}\mathcal{C}(\boldsymbol{z}(t), \boldsymbol{\mu}(t)) \text{ and } u_h(\boldsymbol{c}_h) \geq u_h(\boldsymbol{c}_h^0) \text{ for every } h\right\}. \tag{3}$$

In words, $A(t)$ is the maximum contraction of the consumption possibility set that allows every agent to be kept at least indifferent to the status-quo. Note that the consumption possibility set may be distorted, in the sense that it is not the Pareto frontier defined by technologies. Furthermore, by construction, aggregate efficiency at the status-quo is equal to one: $A(0) = 1$. We say that *outcomes are interior* if the solution to the optimization problem in (3) features strict indifference of every agent.

Real Output. An important statistic is the chain-weighted index of real output or real income. As in the growth accounting literature (and the national accounts), we use the Divisia index to define real output in our model.[20] The real output index at t is defined by

$$\log Y(t) = \int_0^t \sum_i \frac{p_i(s)c_i(s)}{\sum_{i'} p_{i'}(s)c_{i'}(s)} \frac{d \log c_i(s)}{ds} ds,$$

where $c_i(s) = \sum_{h \in H} c_{hi}(s)$ denotes aggregate consumption of good i at $s \in [0, t]$. By construction, real output in the status-quo is one: $Y(0) = 1$.[21]

[20] In principle, using the Arrow-Debreu formalism of indexing commodities by dates and states, $Y(t)$ would correspond to a net-present value type object that incorporates all final demand in every date and state. In practice, statistical agencies apply this formula in a static way, period-by-period, to define real output and use chain-weighted discretized approximations to the true (integral) Divisia index.

[21] As we show below, and is well-understood, chain-weighted real output does not necessarily measure anything welfare relevant. See, for example, Hulten (1973) and, more recently, Baqaee and Burstein (2023). Nevertheless, $\Delta \log Y$ is a useful, and commonly relied upon, statistic which, under the assumptions of

Kaldor-Hicks/Cost-Benefit Efficiency. Another popular aggregate measure is the Kaldor-Hicks efficiency measure. This measure compares the sum of compensating variations to aggregate income at t. If the sum of compensating variations is less than aggregate income, then the winners can hypothetically compensate the losers and there can still be money left-over. The amount of money left over is a measure of the increase in efficiency. This method is the foundation of almost all of cost-benefit style analyses in applied welfare economics and program evaluation in public finance and industrial organization.

Let $e_h(\boldsymbol{p}, u_h)$ be an expenditure function representing preferences $\succeq_h$. The Kaldor-Hicks measure of efficiency at t is

$$A^{KH}(t) = \frac{\sum_h e_h(\boldsymbol{p}(t), u_h(t))}{\sum_h e_h(\boldsymbol{p}(t), u_h(0))}. \tag{4}$$

Note that, by construction, Kaldor-Hicks efficiency at the status-quo is equal to one: $A(0) = 1$.

Consumption-equivalent of Representative Agent. Another well-known aggregate measure, when a representative agent exists, is the consumption-equivalent variation used by Lucas (1987). A *representative agent* is a hypothetical single consumer such that the demand of the representative agent for each good, given prices and total income, coincides with equilibrium quantity of that good, given the same prices and aggregate income.[22]

If a representative agent exists, define the consumption-equivalent for the representative agent, $A^{RA}(t)$, to be

$$u^{RA}\left(\boldsymbol{c}^{RA}(t)/A^{RA}(t)\right) = u^{RA}\left(\boldsymbol{c}^{RA}(0)\right),$$

where u^{RA} is the utility function of the representative agent. In words, $A^{RA}(t)$ is the amount by which the aggregate consumption bundle in t must be contracted to make the positive representative agent exactly indifferent to the status-quo.

3.3 Characterizing $A(t)$ Using Theorem 1

We characterize aggregate efficiency via Theorem 1. To do so, we define a compensated equilibrium, which is a useful fictional construct for proving results and constructing

Proposition 1, is a welfare-relevant measure.

[22] For a formal definition, see Appendix A.

sufficient statistics formulas.[23]

Definition 5 (Compensated Equilibrium)**.** A *compensated equilibrium* is the general equilibrium of an economy with the same technologies, resource constraints, and wedges as the original economy but where there is a representative agent with preferences as in Definition 3. For any equilibrium variable $X(t)$, denote the same variable in the compensated equilibrium by $X^{\text{comp}}(t)$.

It is important to note that compensated equilibrium prices and quantities are not of direct interest themselves, but are instead a useful stepping-stone to calculating changes in aggregate efficiency.

The following result, which is a consequence of Theorem 1, shows that aggregate efficiency can easily be calculated using the compensated equilibrium.

Theorem 2 (Aggregate Efficiency Using Compensated Equilibrium)**.** *If outcomes are interior, then aggregate efficiency can be calculated using the compensated equilibrium:*

$$A(t) = U(\boldsymbol{c}^{comp}(t)) = Y^{comp}(t) = A^{KH,comp}(t) = A^{RA,comp}(t).$$

In words, aggregate efficiency, $A(t)$, can be computed by solving for changes in the utility of the Hicksian representative agent. Furthermore, since there is only one agent in the compensated economy with homothetic preferences, changes in utility coincide with changes in real output in the compensated equilibrium. For the same reason, changes in utility also coincide with Kaldor-Hicks efficiency and consumption-equivalent of the Hicksian representative agent.

The importance of Theorem 2 lies in the fact that it allows every tool and result used to calculate welfare in homothetic representative agent economies to be used to calculate aggregate efficiency with heterogeneous (and non-homothetic) preferences.

Theorem 2 is expressed in terms of endogenous variables in the compensated equilibrium. Solving that equilibrium is essentially the same as solving a representative–agent model, which is well-understood. For readability, we postpone the full characterization of variables in the compensated equilibrium to the end of the paper in Section 7.[24] However, we do note the following useful fact about the compensated equilibrium in this section.

[23] This notion of the compensated equilibrium has many antecedents in prior work, for example it nests the concept in Jones (2002) and Johansson et al. (2022), the Hicksian equilibrium in Baqaee and Burstein (2023), the synthetic equilibrium in Debreu (1951), and is closely related to the adjusted price function in Luenberger (1996). A major difference relative to these notions is that our compensated equilibrium need not be efficient.

[24] In Appendix B we also provide the expenditure function of the Hicksian representative agent in the compensated equilibrium, since the expenditure function is a useful way to solve general equilibrium models.

Lemma 1 (Compensated equilibrium at the status quo). *At the status-quo $t = 0$, prices and quantities in the compensated equilibrium coincide with those in the decentralized equilibrium.*

This lemma, which guarantees that compensated and decentralized variables coincide at the status-quo, is important for calibrating the model when solving for the compensated equilibrium — expenditure shares at the compensated equilibrium in status-quo must coincide with observations.

Before deploying Theorem 2 to construct heterogeneous-agent generalizations of well-known results, we first point out an important, but highly restrictive, special case, where our measure of aggregate efficiency coincides with the other popular alternatives.

3.4 A Miraculous Consensus

To set aside household heterogeneity, a standard benchmark is to assume every household has identical homothetic preferences and faces the same relative prices (i.e. there no household-specific wedges). Under this condition we obtain the following.

Proposition 1 (Miraculous Consensus). *If households have identical homothetic preferences, and face the same relative prices, then a positive representative agent exists and*

$$A(t) = Y(t) = A^{KH}(t) = A^{RA}(t).$$

The first equality can break down if lump-sum transfers are not available.

In words, the change in aggregate efficiency, measured by the consumption-equivalent variation, matches the change in chain-weighted index of real output, Kaldor-Hicks (cost-benefit) efficiency, and the consumption-equivalent of the positive representative agent all in the decentralized equilibrium. That is, under these assumptions, one can compute $A(t)$ without relying on the compensated equilibrium. Each of the underlying assumptions is individually essential: relaxing any one breaks the equivalence. The remainder of the paper explores those more general settings and illustrates, through examples, why our efficiency measure avoids the paradoxes of the alternatives when consensus fails. We consider non-identical and non-homothetic preferences in Section 4, allow for household-specific wedges (and hence prices) in Section 5, and consider limits to redistribution (e.g. no lump-sum taxes) in Section 6.[25]

[25] In this paper, we are focused on household heterogeneity, but even with a single household, the miraculous consensus breaks down when preferences are non-homothetic. This point is discussed in detail by Baqaee and Burstein (2023). Intuitively, when preferences are non-homothetic, even for a single agent, scaling the production possibility set ($A(t)$), the budget constraint ($A^{KH}(t)$), and the equilibrium consumption

Proposition 1 is a consequence of Theorem 2. The reason is that, under the stated assumptions, the price and quantity of each good in the compensated equilibrium coincides with those in the decentralized equilibrium. This implies that $A(t) = Y^{\text{comp}}(t) = Y(t)$, since real output in the compensated and decentralized equilibrium only depend on the prices and quantities of goods. The remaining two equalities are standard.[26]

4 Competitive Economies with Lump-Sum Transfers

In this section, we characterize how aggregate efficiency responds to changes in productivity when the welfare theorems hold. This means that, for the remainder of this section, we assume that all wedges, $\boldsymbol{\mu}(t)$, are all equal to one and lump-sum transfers are available. We consider distorted economies in Section 5 and economies without lump-sum transfers in Section 6.

We begin this section by providing some general comparative static results. We then apply these results to some analytical examples to build intuition — including one example that extends the Arkolakis et al. (2012) (ACR) formula for the gains from trade to allow for non-homothetic preferences and heterogeneous agents. We end this section by comparing our measure of efficiency to measures of efficiency based on the sum of compensating variations (i.e. Kaldor-Hicks efficiency).

4.1 Comparative Statics

Denote the *Domar* weight of each producer or factor i by

$$\lambda_i(t) = \frac{p_i(t)y_i(t)}{\sum_{i'} p_i(t)c_i(t)}\mathbf{1}\{i \text{ is a producer}\} + \frac{w_i(t)z_i(t)}{\sum_{i'} p_i(t)c_i(t)}\mathbf{1}\{i \text{ is a factor}\}.$$

This the sales of i divided by total final expenditures. The following is a well-known result characterizing changes in real output in the decentralized equilibrium.

allocation $(A^{RA}(t))$ do not coincide with one another since, as we shrink resources, the household would want to change the bundle of goods they consume.

[26]Under the stated assumptions, there is a positive representative agent with homothetic preferences, and it follows from Shephard's lemma that $Y(t) = A^{RA}(t)$ (see, e.g., Baqaee and Burstein, 2023). The final equality in Proposition 1 follows from the fact that the indirect utility function of each agent, $v(\boldsymbol{p}, I_h)$ can be written as $I_h/P(\boldsymbol{p})$, where $P(\boldsymbol{p})$ is an ideal price index. It then follows that, $v^{RA}(\boldsymbol{p}, \sum_h I_h)$, can be written as $(\sum_h I_h)/P^{RA}(\boldsymbol{p})$. Hence, $A^{RA}(t) = v^{RA}(\boldsymbol{p}(t), \sum_h I_h(t))/v^{RA}(\boldsymbol{p}(0), \sum_h I_h(0)) = \sum_h I_h(t)/(\sum_h I_h(0)) \times P(\boldsymbol{p}(0))/P(\boldsymbol{p}(t)) = A^{KH}(t)$.

Proposition 2 (Hulten's Theorem)**.** *The change in chain-weighted real output is*

$$\Delta \log Y = \int_0^t \sum_i \lambda_i(s) \frac{d \log z_i}{ds} ds. \tag{5}$$

This formula, which generalizes Solow (1957), shows that the elasticity of real output to the productivity of producer i or the quantity of factor i is just the Domar weight of i. Using Theorem 2, we can easily state a version of Hulten's theorem that applies to our measure of aggregate efficiency instead.

Proposition 3 (Compensated Hulten's Theorem)**.** *The change in aggregate efficiency at t is*

$$\Delta \log A = \int_0^t \sum_i \lambda_i^{comp}(s) \frac{d \log z_i}{ds} ds. \tag{6}$$

In Section 7, we characterize $\lambda_i^{comp}(s)$ explicitly as a function of the productivity changes $\Delta \log z$, elasticities of substitution, and expenditure shares.

Differentiating (6) with respect to t and evaluating at $t = 0$ shows that, to a first-order approximation, the change in aggregate efficiency, $\Delta \log A$, coincides with the change in real output in the competitive equilibrium $\Delta \log Y$.

Corollary 1 (First Order Changes in Aggregate Efficiency)**.** *To a first-order approximation, the change in aggregate efficiency is*

$$\Delta \log A \approx \sum_i \lambda_i^{comp}(0) \Delta \log z_i = \sum_i \lambda_i(0) \Delta \log z_i \approx \Delta \log Y.$$

The second equality follows from Lemma 1, which states that prices and quantities in the compensated equilibrium $t = 0$ are equal to those in decentralized economy. In other words, the first-order version of Hulten's theorem applies unaltered to aggregate efficiency.

Baqaee and Farhi (2019c) show that, to a second-order approximation, changes in real output are given by

$$\Delta \log Y \approx \sum_i \lambda_i \Delta \log z_i + \frac{1}{2} \sum_i \Delta \lambda_i \Delta \log z_i.$$

Differentiating (6) twice with respect to t and evaluating at $t = 0$, gives the following generalization of Baqaee and Farhi (2019c).

Proposition 4 (Second Order Changes in Aggregate Efficiency). *To a second-order approximation, the change in aggregate efficiency due to changes in primitives is*

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

where λ_i and $\Delta\lambda_i^{comp}$ are evaluated at status-quo. In Section 7, we write $\Delta\lambda_i^{comp}$ explicitly as a function of the productivity changes $\Delta \log z$, microeconomic elasticities of substitution, and expenditure shares in the status-quo.

Proposition 4 shows that discrepancies between aggregate efficiency $\Delta \log A$ and real output $\Delta \log Y$ start at the second-order, since, generically $\Delta\lambda \neq \Delta\lambda^{\text{comp}}$. We delay an explicit formula for $\Delta\lambda^{\text{comp}}$ to Section 7. The intuition is that $\Delta\lambda^{\text{comp}}$ is simply the change in Domar weights in a special case of the environment considered by Baqaee and Farhi (2019c) where the consumption growth of each agent is treated as-if it is a final good, and there is a Leontief final demand aggregator over final goods.

4.2 Analytical Examples

To build some intuitioin, we work through some analytical examples for how $\Delta \log A$ responds to changes in technologies.

Example 3 (Regional Productivity Shocks). Consider households in different regions, indexed by h, with preferences over tradeable goods and locally produced nontradeable services:

$$u_h(\boldsymbol{c}) = c_g^{\alpha} c_s^{1-\alpha}, \qquad \sum_h c_{hg} = z_g, \qquad c_{hs} = z_{hs}.$$

The first equation shows that utility in each region depends on goods and services with the expenditure share on goods equal to α. The second equation is the resource constraint for goods, which clear at the aggregate level. The third equation is the resource constraint for services, which clear region-by-region, since services are not traded. The parameters z_g and z_{hs} control the endowments of goods and services.

Suppose that households in region h own the local endowment of services and own a share χ_h of the aggregate endowment of the traded good. This implies that, in equilibrium, χ_h is the expenditures of each household as a share of total consumption expenditures. The Domar weight on goods is $\lambda_g = \sum_h \chi_h \alpha = \alpha$ and the Domar weight on services in region h is $\lambda_{hs} = \chi_h(1-\alpha)$.

When productivities change, according to Proposition 2, the change in real output is

$$\Delta \log Y = \alpha \Delta \log z_g + (1-\alpha)\mathbb{E}_{\chi}\left[\Delta \log z_s\right],$$

where $\mathbb{E}_{\chi}\left[\Delta \log z_s\right]$ is the average productivity shock to services weighted by the vector χ. This expression is exact because, in the competitive equilibrium, $\Delta\lambda = 0$. Furthermore, since the Domar weights are constant in the competitive equilibrium, there is a positive representative agent with Cobb-Douglas preferences over goods and services in all regions. Since the positive representative agent has homothetic preferences, the change in real output $\Delta \log Y$ coincides with the change in the welfare of this positive representative agent $\Delta \log A^{RA}$.

According to Proposition 4, the change in aggregate efficiency, to a second-order, is

$$\Delta \log A \approx \alpha \Delta \log z_g + (1-\alpha)\mathbb{E}_{\chi}\left[\Delta \log z_s\right] - \frac{1}{2}\frac{(1-\alpha)^2}{\alpha} Var_{\chi}\left[\Delta \log z_s\right] \le \Delta \log Y = \Delta \log A^{RA}.$$

The miraculous consensus of Proposition 1 fails because the agents do not have the same preferences. As predicted by Corollary 1, $\Delta \log Y$ and $\Delta \log A$ do coincide to a first-order, since the discrepancy scales in the square of $\Delta \log z_s$. The second-order approximation shows that $\Delta \log A$ is a concave envelope of $\Delta \log Y$ around the status-quo $\Delta \log \boldsymbol{z} = 0$ — amplifying negative shocks and mitigating positive shocks to services relative to real output. Intuitively, negative shocks to services are "more costly" since the losses cannot be shared across regions, whereas the positive representative agent is "willing" to substitute between regions.

Our next example uses Theorem 1 to apply a version of the popular Arkolakis et al. (2012) (ACR) formula to economies with heterogeneous agents with non-homothetic preferences.

Example 4 (Gains from Trade in Armington with Heterogeneous Non-Homothetic Agents). Consider a country with different consumers, h, that value domestic and foreign goods differently:

$$u_h(\boldsymbol{c}_h) = \left[(\alpha_h)^{\frac{1}{\theta_h}} \left(u_h(\boldsymbol{c}_h)\right)^{\zeta_h} c_{hd}^{\frac{\theta_h - 1}{\theta_h}} + (1-\alpha_h)^{\frac{1}{\theta_h}} c_{hf}^{\frac{\theta_h - 1}{\theta_h}} \right]^{\frac{\theta_h}{\theta_h - 1}}.$$

The parameter α_h controls home bias for household h, $\theta_h > 1$ is the compensated Armington elasticity, and the parameter ζ_h controls the degree of non-homotheticity for agent h. We consider the gains from trade relative to autarky, by raising iceberg trade costs to infinity.

The country trades with the rest of the world in the status-quo. The efficiency gain from trade relative to autarky, $\Delta \log A$, is the minimal increase in the domestic good needed in autarky to keep every consumer indifferent to the status-quo.

Let s_{hd}^0 denote household h's budget share on the domestic good in the status-quo. Replicating the ACR argument for the Hicksian representative agent in the compensated equilibrium, we can show that losses for this country of going to autarky are given by

$$\Delta \log A = -\log \mathbb{E}_{\chi^0}\left[\left(s_{hd}^0\right)^{\frac{-1}{\theta_h - 1}}\right].$$

That is, calculate an ACR formula for each household and then average these using household expenditures in the status-quo (denoted by χ^0). Interestingly, the non-homotheticity plays no role whatsoever and the utility elasticities ξ_i are not needed to calculate $\Delta \log A$. In particular, if there is a single agent, the equation above shows that ACR holds without change even if preferences are non-homothetic as long as we use the compensated Armington trade elasticity.[27]

To get more intuition of how household heterogeneity matters, consider a second-order approximation of $\Delta \log A$ around autarky:[28]

$$\Delta \log A \approx \mathbb{E}_\chi\left[\frac{\log s_{hd}^0}{\theta_h - 1}\right] - \frac{1}{2} Var_\chi\left[\frac{\log s_{hd}^0}{\theta_h - 1}\right]. \tag{7}$$

The first term is just an "average" version of the ACR formula — the ACR formula is applied household-by-household and then averaged using households' share of aggregate income χ_h. The second term is the Jensen's inequality term, and it lowers aggregate efficiency if there is any heterogeneity in households' exposure to traded goods either due to variance in expenditure shares, s_{hd}^0, or trade elasticities, θ_h.

4.3 Comparison to Kaldor-Hicks efficiency

We end this section by comparing our measure of efficiency with the popular Kaldor-Hicks measure. As we discuss in Section 6, a major difference between our approach and Kaldor-Hicks is that our measure does not assume lump-sum taxes are available.

[27] If preferences are non-homothetic, then there is a distinction between the compensated and uncompensated trade elasticities. If we have estimates of the latter, one must use Slutsky's equation to first convert them into the compensated elasticities (see, e.g. 2024 Auer et al.).

[28] This is an approximation in $\frac{\log s_{hd}^0}{\theta_h - 1}$ around $s_{hd}^0 = 1$. To derive this, we follow the strategy in Theorem 3 of Baqaee and Farhi (2019a) who consider the gains from trade with a homothetic representative agent.

However, even when lump-sum taxes are available, as in this section, our measure still does not necessarily coincide with Kaldor-Hicks efficiency. The following proposition illustrates this fact.

Proposition 5 (Paradox for Kaldor-Hicks Efficiency)**.** *For any change in technologies (movements of the Pareto efficient frontier), the change in aggregate efficiency, measured using aggregate consumption-equivalent variation, is weakly less than Kaldor-Hicks efficiency:*

$$\Delta \log A \leq \Delta \log A^{KH}.$$

For pure redistributions (movements along the Pareto efficient frontier), the change in aggregate efficiency, measured using aggregate consumption-equivalent variation, is zero, whereas the change in Kaldor-Hicks efficiency can be positive:

$$\Delta \log A = 0 \leq \Delta \log A^{KH}.$$

These inequalities are strict outside of knife-edge cases. The final inequality is restatement of the Boadway (1974) paradox — the observation that the sum of compensating variations assigns strictly positive value to pure redistributions in general equilibrium when relative prices respond to transfers.[29]

Figure 4 graphically illustrates the Boadway paradox using a two-good, two-consumer economy. Intuitively, redistributions lower the relative price of those goods that are more valued by the losers. Hence, in the new equilibrium, it is relatively cheap to compensate these households. Of course, such compensations are, in practice, infeasible because if they were to occur, then relative prices would rise for those households that need compensation.[30] In fact, it is possible to construct examples where the production possibility set of the economy shrinks, $\mathcal{C}(\boldsymbol{z}(t)) \subset \mathcal{C}(\boldsymbol{z}(0))$, so that $\Delta \log A < 0$ and where $\Delta \log A^{KH} > 0$. This happens if the competitive equilibrium associated with $\mathcal{C}(\boldsymbol{z}(t))$ has very different relative prices to $\mathcal{C}(\boldsymbol{z}(0))$, so that it is relatively cheap for the winners to compensate the losers under the new prevailing prices.

There is a special case where $A^{KH}(t)$ and $A(t)$ coincide: when relative prices do not depend on final demand. This happens when the economy's production possibility frontier is a hyperplane, so the marginal rate of transformation, and hence relative prices, are determined by technology.[31] Under these conditions, the change in aggregate efficiency,

[29]See also Blackorby and Donaldson (1990) for a related critique of the sum of compensating variations as a measure of efficiency.

[30]See Jones (2002) for a detailed discussion.

[31]See Proposition 12 in Appendix C for a formal statement.

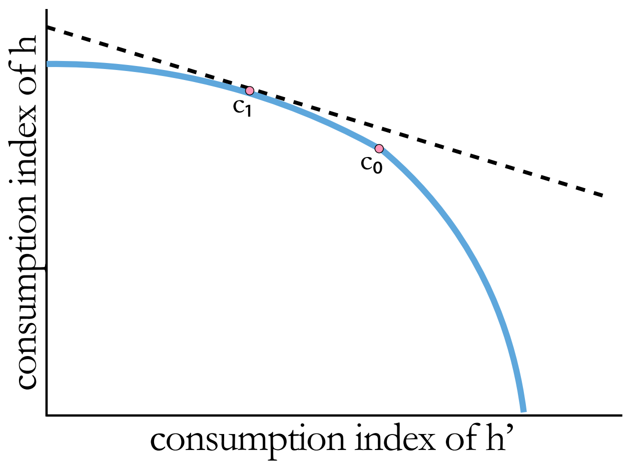


Figure 4: The sum of compensating variations is less than aggregate income since c^0 is below the dashed straight line.

$\Delta \log A$, coincides with the change in Kaldor-Hicks efficiency as defined in Equation (4). Given our assumption that production functions are all constant-returns to scale, the production possibility frontier becomes a hyperplane whenever there is only one primary factor of production.

Besides the endogeneity of prices, there is another important reason why our measure of efficiency can differ from the Kaldor-Hicks measure. The Kaldor-Hicks measure, by summing up compensating variations, implicitly assumes that lump-sum transfers are available, so that winners can costlessly compensate the losers (assuming relative prices are constant). Our definition of aggregate efficiency naturally extends to allow for limited redistribution, as we discuss further in Section 6.

5 Distorted Economies with Lump-Sum Transfers

We now relax the assumption in the previous section that there are no wedges, and we characterize how aggregate efficiency responds to changes in productivities and wedges. One important comparative static we focus on is the efficiency losses from misallocation — the increase in aggregate efficiency caused by the elimination of all wedges.

Notation. Throughout this section we normalize household-good-specific wedges to one, $\mu_{hi} = 1$. This involves no loss of generality: any distortion to household h's consumption of good i can be represented by inserting a fictitious intermediary that purchases good i on behalf of h and placing an output wedge on that intermediary. We make this normalization to simplify notation.

5.1 Comparative Statics for Changes in Technologies and Wedges

Theorem 2 means that we can convert results about real output into results about aggregate consumption-equivalent variation by applying them to variables in the compensated equilibrium. For example, consider the following generalization of Petrin and Levinsohn (2012) and Baqaee and Farhi (2019b).

Proposition 6 (Changes in Aggregate Efficiency with Wedges)**.** *In response to changes in wedges and productivities, the change in aggregate efficiency is*

$$\Delta \log A = \int_0^t \sum_i \lambda_i^{comp}(s) \left[\left(1 - \frac{1}{\mu_i(s)}\right) \frac{d \log y_i^{comp}}{ds} + \frac{1}{\mu_i(s)} \frac{d \log z_i}{ds} \right] ds.$$

In Section 7, we characterize $\lambda_i^{comp}(s)$ and $d \log y_i^{comp} / ds$ explicitly as a function of the productivity changes $\Delta \log z$, elasticities of substitution, and expenditure shares.

The compensated Domar weights, λ^{comp}, and quantities, $d \log y^{\text{comp}}$, can be computed using standard methods for inefficient economies with homothetic representative agents (see Section 7).

We can contrast Proposition 6 with Harberger's social welfare formula. In his classic paper, Harberger (1971) argued that the welfare effect of a policy that changes quantities $\{y_i\}$ over time should be computed as

$$\int_0^t \sum_i [p_i(s) - mc_i(s)] \frac{dy_i}{ds} \, ds \;=\; \int_0^t \sum_i \lambda_i(s) \left(1 - \frac{1}{\mu_i(s)}\right) \frac{d \log y_i}{ds} \, ds, \tag{8}$$

where the equality uses the fact that final expenditure is the numeraire. In words, he argued that whenever a good's marginal benefit, $p_i(s)$, exceeds its marginal cost, $mc_i(s)$, then expanding its quantity (holding others fixed) raises "social welfare." Proposition 6 shows that this expression exactly measures the change in aggregate efficiency when (i) households have identical homothetic preferences, (ii) no household–good–specific wedges exist, and (iii) lump–sum transfers are available; under these conditions the compensated Domar weight λ^{comp} and changes in quantities $d \log y^{\text{comp}}$ coincide with their uncompensated counterparts in the decentralized equilibrium and Harberger's formula holds. Outside of these cases, we must use the compensated versions of the objects. In

Section 6, we show that a version of Proposition 6 holds even without lump-sum transfers.

5.2 Misallocation and the Distance to Pareto Frontier

We now focus on a particular counterfactual: we apply Proposition 6 to compute the economic waste caused by distortions. Let $\boldsymbol{\mu}$ be a vector of wedges. We measure economic waste by how far the Pareto frontier can be contracted while keeping every agent at least as well off as under the status quo.

Formally, denote the status–quo allocation by $\boldsymbol{c}^0(\boldsymbol{\mu})$ and the consumption possibility set by $\mathcal{C}(\boldsymbol{\mu})$ (we suppress productivity parameters since we hold them fixed). By the second welfare theorem, the Pareto frontier is $\mathcal{C}(\mathbf{1})$. The efficiency gains of eliminating markups are measured by

$$A\big(\boldsymbol{c}^0(\boldsymbol{\mu}),\, \mathcal{C}(\mathbf{1})\big).$$

With a complete structural model this term can be computed using Theorem 2. However, below we derive an approximation that is more intuitive and requires less information to be applied.[32]

Proposition 7 (Harberger Triangles)**.** *To a second-order approximation in* $\log \mu$*, the change in aggregate efficiency is*

$$\Delta \log A \approx -\frac{1}{2} \sum_i \lambda_i d \log y_i^{comp} \log \mu_i, \tag{9}$$

where $d \log y_i^{comp} \equiv \sum_j \frac{\partial \log y_i^{comp}}{\partial \log \mu_j} \log \mu_j$ *is the change in the quantity of i caused by the wedges in the compensated equilibrium. The approximation error is order* $\log \mu^3$*. The derivatives and expenditure shares in* (9) *are evaluated at the status-quo.*

This proposition generalizes deadweight loss triangles to measure aggregate efficiency losses from wedges in general equilibrium economies with heterogeneous agents and non-homothetic preferences. The proof relies on translating results from Baqaee and Farhi (2024) using Theorem 2.

There are two advantages to using Proposition 7 over and above simply applying Proposition 6 using a fully-spelled out structural model. First, the Harberger triangles formula can be used to get analytical intuition for misallocation costs through the use of loglinearized expressions, as demonstrated below. Second, as we show in a companion

[32] Baqaee and Burstein (2025b) is a fleshed out application of Proposition 7 to estimate misallocation due to financial market incompleteness without assuming a fully structural model.

paper, Baqaee and Burstein (2025b), it is possible to populate the terms in (7) with considerably fewer assumptions about the primitives of the economy — e.g. the drivers of distortions, productivity processes, and so on.

The intuition for (9) is familiar — a wedge on i is more costly the higher is the Domar weight and the more elastic is the quantity of i relative to the wedge. However, compared to a representative agent model with homothetic preferences, the relevant notion of elasticity here is the one in the compensated equilibrium, not the decentralized one.

5.3 Analytical Examples

We provide some pen-and-paper examples to build intuition.

Example 5 (Misallocation when Markups Vary by Household). Suppose each agent h's has CES preferences over consumption goods with elasticity of substitution θ_h and h's share of spending in the status-quo is given by the vector $\boldsymbol{b}_h$. We consider a situation in which each household h pays potentially a different markup μ_{hi} on each good i.[33] Suppose that all consumption goods are ultimately produced linearly from a single common primary factor called labor, which is inelastically supplied.

We can apply Proposition 7 to write the aggregate efficiency losses, up to a second order approximation as[34]

$$\Delta \log A \approx \frac{1}{2}\mathbb{E}_{\chi}\left[\theta_h Var_{\boldsymbol{b}_h}\left[\log \boldsymbol{\mu}_h | h\right]\right] = \frac{1}{2}\sum_h \chi_h \theta_h \sum_n b_{hn}\left[\log \mu_{hn} - \sum_{n'} b_{hn'} \log \mu_{hn'}\right]^2, \quad (10)$$

In words, the reduction in efficiency caused by the markups depends on the average variance in markups paid by each household multiplied by that household's elasticity of substitution. Intuitively, if θ_h is very high, then dispersion in markups faced by h causes a greater reduction in aggregate efficiency. Furthermore, aggregate efficiency falls by more if richer households (those with higher χ_h) are exposed to more markup dispersion. Importantly, this expression does not depend on the *level* of markups paid by each household. A proportional scaling of all markups paid by any household would leave this expression unchanged because increasing all markups on a single household is equivalent

[33] Formally, hi indexes the intermediary between good i and household h. We assume that this intermediary charges a markup of μ_{hi} on its marginal cost. The intermediary's marginal cost is just the price of good i.

[34] This formula is a special case of Proposition 10 in Section 7, where we explicitly solve for prices and quantities in the compensated equilibrium as a function of primitives.

to a lump-sum tax on that household, and has no effect on aggregate efficiency.[35]

We compare $\Delta \log A$ to changes in real output, $\Delta \log Y$, and the welfare of a positive representative agent, $\Delta \log A^{RA}$, as markups are eliminated.

Example 6 (Real Output and Positive Representative Agent Losses from Markups). The change in chain-weighted real output, and the existence of a positive representative agent, depends on how wedge revenues are distributed between households.[36] Suppose that as we eliminate markups, each households' share of income, χ_h, stays constant.

Since the distribution of income is constant, there is a normative representative agent with Cobb-Douglas preferences across each households' consumption bundle (i.e. an agent whose utility is maximized by observed allocations). The change in the welfare of this representative agent, in consumption-equivalent terms, is equal to the change in chain-weighted real output, and both are equal to a second-order approximation to

$$\Delta \log Y = \Delta \log A^{RA} \approx \Delta \log A + \frac{1}{2} Var_{\boldsymbol{\chi}} \left[\mathbb{E}_{\boldsymbol{b}_h} \left[\log \boldsymbol{\mu}_h | h \right] \right].$$

Since the last summand on the right-hand side is always non-negative, the change in real output and the welfare of the representative agent are strictly larger than the change in aggregate efficiency. One limiting case of this is where all markup-variation is at the household level. In this case, $\Delta \log A = 0$ because the economy is already on the efficient frontier and eliminating markups is purely redistributive. However, in this example, $\Delta \log Y = \Delta \log A^{RA} > 0$.

We can push this example even farther: suppose that all markup-variation is at the household level, and as we eliminate all markups, we also change the productivity of labor by $\Delta \log z < 0$ at the same time. In this case, the change in chain-weighted real output is

$$\Delta \log Y \approx \Delta \log z + \frac{1}{2} Var_{\boldsymbol{\chi}} \left[\mathbb{E}_{\boldsymbol{b}_h} \left[\log \boldsymbol{\mu}_h | h \right] \right],$$

[35] This is also true for the fully non-linear solution of aggregate efficiency. To contrast our approach, in (10), to the formulas derived by Hsieh and Klenow (2009) or Baqaee and Farhi (2020) for misallocation in representative agent models, consider the following alternative model. There is a single household that has CES preferences over c_{hn} with elasticity of substitution θ. In this case, the losses in efficiency from markups, up to a second order are: $\frac{1}{2}\theta Var_{\boldsymbol{\lambda}} [\log \mu]$, where $Var_{\boldsymbol{\lambda}}(\cdot)$ denotes a variance with $\lambda_{hn} = \chi_h b_{hn}$ as the weights. If $\theta_h = \theta$, then we can rewrite this as $\frac{1}{2}\theta Var_{\boldsymbol{\lambda}} [\log \mu] = \Delta \log A + \frac{1}{2}\theta Var_{\boldsymbol{\chi}} \left[\mathbb{E}_{\boldsymbol{b}_h} \left[\log \boldsymbol{\mu}_h | h \right] \right]$, which is an application of the law of total variance, which shows that the losses are strictly lower in the heterogeneous agent model.

[36] In fact, it is possible to construct a path whereby real output falls even though we eliminate wedges. This happens because, outside of the homothetic representative agent case, real output is generically path-dependent and not related to any welfare measure (see, e.g., Hulten, 1973; Baqaee and Burstein, 2023).

whereas $\Delta \log A = \Delta \log z$. Hence, if the physical productivity shock is small enough, then real output rises, even though the Pareto frontier shifted inwards.

To summarize, in this example, $\Delta log Y$ and $\Delta \log A^{RA}$ assign a positive value to a pure transfer, and if negative productivity shocks are sufficiently small, then a positive value to a strictly smaller production possibility set.

---

6 Aggregate Efficiency with Limits to Redistribution

In this section, we extend the analyses in Sections 4 and 5 to allow for imperfect redistributive tools. This is another advantage of our approach relative to measures based on adding up willingness-to-pay across all households (e.g. as in Kaldor-Hicks). Intuitively, when we add up willingness-to-pay, we implicitly assume that winners can compensate losers. In Section 4, we illustrated one issue with this approach: monetary compensations can change relative prices so that, in practice, the necessary compensations are infeasible. In this section, we focus on a second issue — monetary compensations may be infeasible because lump-sum transfers are not available.

Theorem 1 applies regardless of what redistributive tools are available. In this section, we apply Theorem 1 to the case where redistribution can only be achieved via linear taxation in general equilibrium with wedges.

6.1 General Solution with Linear Taxes

Consider a general equilibrium with technologies $\boldsymbol{z}$ and wedges $\boldsymbol{\mu}$. We allow a vector of linear taxes $\boldsymbol{\tau}$ on different goods, and let the vector $\boldsymbol{T}$ dictate the amount of tax revenues sent to each household. We require budget-balance, so that total tax revenues must equal total transfers to households. Index the equilibrium consumption allocation $\boldsymbol{c}(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{T})$ by technologies, $\boldsymbol{z}$, wedges, $\boldsymbol{\mu}$, and the tax-and-transfer scheme, $(\boldsymbol{\tau}, \boldsymbol{T})$.

Let the set of all feasible tax-and-transfer schemes be $\mathcal{T}(\boldsymbol{z}, \boldsymbol{\mu})$. The case with lump-sum transfers studied in Section 4 and Section 5 is the special case that places no non-negativity constraint on the vector $\boldsymbol{T}$ so that redistribution is accomplished without linear taxes. In this section, we allow for the possibility that this set has other restrictions. For example, the feasible set $\mathcal{T}$, may allow only distortionary linear taxes on some subset of goods, limit lump-sum transfers to be non-negative.

Corollary 2 (Aggregate Efficiency with Restricted Tax-and-Transfer Instruments)**.** *Theo-*

rem 1 implies that aggregate efficiency satisfies:[37]

$$A(\boldsymbol{z}, \boldsymbol{\mu}) = \max \left\{ U\left(\boldsymbol{c}(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{T})\right) : (\boldsymbol{\tau}, \boldsymbol{T}) \in \mathcal{T}(\boldsymbol{z}, \boldsymbol{\mu}) \right\}. \tag{11}$$

That is, aggregate efficiency is given by the highest utility $U(\boldsymbol{c})$ that can be achieved by choosing feasible taxes and transfers, $(\boldsymbol{\tau}, \boldsymbol{T}) \in \mathcal{T}(\boldsymbol{z}, \boldsymbol{\mu})$, taking into account how those choices affect consumption, $\boldsymbol{c}$, and in turn utility, U.

In words, $A(\boldsymbol{z}, \boldsymbol{\mu})$ measures the maximum contraction (or minimum expansion) of the set of feasible equilibrium consumption allocations, given technologies, $\boldsymbol{z}$, wedges, $\boldsymbol{\mu}$, and tax-and-transfer instruments, $(\boldsymbol{\tau}, \boldsymbol{T})$. Corollary 2 applies to the special cases considered in Section 4 and 5, where we assume that the feasible set of instruments, $\mathcal{T}$, consists only of unrestricted lump-sum transfers.

As before, we index technologies and wedges by a scalar t and let $t = 0$ denote the status-quo. Figure 5 illustrates $A(t)$ using a two household example. In the figure, the status-quo allocation, $\boldsymbol{c}(0)$, and the decentralized allocation without transfers, $\boldsymbol{c}(t)$, are denoted by red circles. The solid blue line shows the feasible consumption possibility frontier at t given distortionary taxation, and the dashed line indicates the frontier at t given unrestricted lump-sum taxes. The two frontiers touch at the decentralized point, since the decentralized point is does not engender any distortionary redistributive taxation. However, the solid blue set is strictly smaller than the dashed line since distortionary taxation limits the set of feasible redistributions. The change in efficiency, $\Delta \log A$, is still the largest radial contraction of $\mathcal{C}(t)$ that allows every household to be made at least indifferent to the status-quo. Since the larger is the possibility set $\mathcal{C}(t)$, the more it must be contracted to reach indifference, aggregate efficiency gains are larger with better redistributive tools.

Let $\boldsymbol{\tau}^*(t)$ and $\boldsymbol{T}^*(t)$ to be the maximizers of (11). Using $\boldsymbol{\tau}^*(t)$, we provide a slightly more general definition of the compensated equilibrium.

Definition 6 (Compensated Equilibrium)**.** A *compensated equilibrium* is the general equilibrium of an economy with the same technologies, resource constraints, wedges, and linear taxes $\boldsymbol{\tau}^*(t)$ but where there is a representative agent with preferences as in Definition 3. For any equilibrium variable $X(t)$, denote the same variable in the compensated equilibrium by $X^{\text{comp}}(t)$.

[37] If there are multiple equilibria, then $\boldsymbol{c}(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{T})$ is a correspondence and the maximization is applied to set of potential equilibrium allocations. We could equivalently write $A(\boldsymbol{z}, \boldsymbol{\mu}) = \max_{\boldsymbol{c} \in \mathcal{C}(\boldsymbol{z}, \boldsymbol{\mu})} U(\boldsymbol{c})$, where the consumption possibility set is $\mathcal{C}(\boldsymbol{z}, \boldsymbol{\mu}) = \{\boldsymbol{c}(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{T}) : (\boldsymbol{\tau}, \boldsymbol{T}) \in \mathcal{T}(\boldsymbol{z}, \boldsymbol{\mu})\}$.

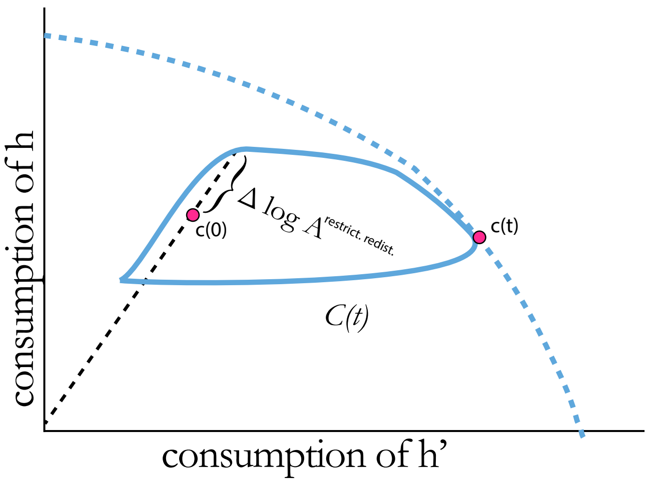


Figure 5: Aggregate efficiency is measured by the maximal radial expansion of the feasible set necessary to achieve indifference.

The following result, which is a consequence of Theorem 1, generalizes Theorem 2 to allow for limited redistribution.

Theorem 3 (Aggregate Efficiency Using Compensated Equilibrium)**.** *If outcomes are interior, then aggregate efficiency can be calculated using the compensated equilibrium:*

$$A(t) = Y^{comp}(t) = A^{KH,comp}(t) = A^{RA,comp}(t).$$

In words, aggregate efficiency, $A(t)$, can be computed by solving for changes in real output, or welfare of the Hicksian representative agent, in the compensated equilibrium.[38] Once again, this means that tools and results used to calculate welfare in homothetic representative agent economies can be converted into results about aggregate efficiency with heterogeneous and non-homothetic preferences.

The main challenge lies in knowing the necessary taxes $\tau^*(t)$ which the proposition takes as given. However, given these taxes, then the change in every price and quantity in the compensated equilibrium can be calculated as a function of t by applying the results in Baqaee and Farhi (2020).

An interesting consequence of Theorem 3 is the following generalization of Proposition 4.

Proposition 8 (Productivity Shocks with Limited Redistribution)**.** *Consider a perfectly competitive status-quo without linear taxes and wedges. If outcomes are interior, the response of*

[38]The fact that Kaldor-Hicks efficiency, $A^{KH,comp}(t)$, in the compensated equilibrium with distortionary taxes τ^* coincides with the rest follows trivially from the fact that the compensated equilibrium has a single agent with homothetic preferences. It is important to note that $A^{KH,comp}(t)$ is not the same as $A^{KH}(t)$.

aggregate productivity to a productivity shock, $\Delta \log \boldsymbol{z}$, to a second-order approximation, is given by

$$\Delta \log A = \sum_i \left(\lambda_i + \frac{1}{2} \sum_j \frac{\partial \lambda_i^{comp}}{\partial \log z_j} \Delta \log z_j \right) \Delta \log z_i + \frac{1}{2} \sum_i \lambda_i \left(\sum_j \frac{\partial \log y_i^{comp}}{\partial \log \tau_j} \Delta \log \tau_j^* \right) \Delta \log \tau_i^*. \tag{12}$$

The first set of summands are exactly as in Proposition 4. The second set of summands, which are new and non-positive, capture the inefficiency caused by imperfect redistribution. These are the sum of Harberger triangles associated with the linear taxes in $\tau^*(t)$. If lump-sum taxation is not feasible, then linear taxes must be used, $\Delta \log \tau^* \neq 0$, and aggregate efficiency with limited redistribution is lower than with lump-sum taxation by exactly the sum of deadweight loss triangles. That is, the response of aggregate efficiency to productivity shocks is the same as it would be if lump-sum transfers were possible minus the deadweight loss triangles associated with distortionary taxes. The simplicity of Equation (12) follows from the fact that the status-quo is undistorted. This ensures that (1) there are no interactions of taxes with pre-existing distortions, (2) the cross-partials between $d \log \tau^*$ and $d \log z$ are all zero due to the envelope theorem. Corollary 2 and Theorem 3 do not require that the status-quo be undistorted.

The following simple corollary, obtained by ignoring the second order terms, shows that Hulten's theorem holds, without change, even with limited redistribution.

Corollary 3 (Hulten's Theorem with Limited Redistribution)**.** *Consider a perfectly competitive status-quo without linear taxes and wedges. If outcomes are interior, the response of aggregate productivity to a productivity shock, $\Delta \log \boldsymbol{z}$, to a first-order approximation, is given by*

$$\Delta \log A = \sum_i \lambda_i \Delta \log z_i.$$

Intuitively, the losses from costly-redistribution are second-order, and hence to a first-order approximation, only the direct effects of the productivity shock matter (assuming we start at a competitive equilibrium).

6.2 Worked Out Examples

We now use two examples to study the efficiency gains from international trade and from skill-biased technological change, accounting for limited redistribution. We use the approximation in Proposition 8 to provide intuition. We check the numerical performance of the second-order approximation by computing exact results using Theorem 3.

Example 7 (Gains from Trade with Limited Redistribution). We revisit Example 4, which studied the gains from trade, but this time we incorporate limits to redistribution. This generalizes Arkolakis et al. (2012) to allow for heterogeneous agents and limited redistribution.

Suppose there are two households, and household h has nested-CES preferences over domestic consumption goods, c_{hd}, imported goods, c_{hf}, and a second domestic good we call leisure l_h:

$$u_h(\boldsymbol{c}_h) = \left[(1-\gamma_h)^{\frac{1}{\rho}} \left[(1-\alpha_h)^{\frac{1}{\theta}} c_{hd}^{\frac{\theta-1}{\theta}} + \alpha_h^{\frac{1}{\theta}} c_{hf}^{\frac{\theta-1}{\theta}} \right]^{\frac{\theta}{\theta-1}\frac{\rho-1}{\rho}} + \gamma_h^{\frac{1}{\rho}} l_h^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}}.$$

The model in Example 4 did not feature the leisure good. The inner nest combines domestic and foreign consumption goods with Armington elasticity θ and home bias controlled by the parameter α_h. The outer nest combines the goods bundle with leisure with elasticity of substitution ρ and share parameter γ_h.[39]

Household h is endowed with one unit of time and a_h efficiency units of labor and faces a budget constraint:

$$\tau p_d c_{hd} + \tau p_f c_{hf} = w a_h (1 - l_h) + T_h,$$

where p_d and p_f denote the price of each consumption good, w_h is the wage per efficiency unit, τ is the gross-tax rate on consumption, and T_h is a lump sum transfer. Budget balance requires $(\tau - 1)\sum(p_d c_{hd} + p_f c_{hf}) = \sum T_h$. The domestic consumption good is produced linearly with labor, so the resource constraint for domestic consumption is $\sum_h c_{hd} = \sum_h a_h(1 - l_h)$, with $p_d = w$. The resource constraint for leisure is $l_h \leq 1$.

The status-quo is a competitive equilibrium without taxes in which the country trades with the rest of the world. We consider the gains from trade relative to autarky, by raising iceberg trade costs to infinity. The efficiency gain from trade relative to autarky, $\Delta \log A$, is the minimal increase in the domestic consumption possibility set in autarky needed to keep every consumer indifferent to the status-quo. The consumption possibility set encodes the potentially distortionary impact of taxes required to transfer income between households. We compare two cases: (1) lump-sum taxation is available and the second welfare theorem holds; (2) lump-sum taxation is not available, $T_h \geq 0$, and linear consumption taxes must be used.

[39] For simplicity of exposition, we abstract from non-homotheticities and differences in elasticity parameters across households. It is simple to extend the model in this way.

Let

$$\Omega_d^0 = \frac{p_d^0 c_{hd}^0 + p_f^0 c_{hf}^0}{w^0 a_h}$$

denote household h's budget share on consumption in the status-quo as a share of the value of h's total time endowment (the remainder is implicit expenditures on leisure). For simplicity of exposition, and since it is fairly realistic, we assume that both households work the same number of hours in the status-quo, which implies that Ω_d^0 does not vary by household. Let s_{hd}^0 denote household h's share of expenditures on the domestic consumption good relative to all consumption goods:

$$s_{hd}^0 = \frac{p_d^0 c_{hd}^0}{p_d^0 c_{hd}^0 + p_f^0 c_{hf}^0}.$$

Lump-Sum Taxation. With lump-sum taxation, using Proposition 8, we can write the gains from trade to a second-order approximation as

$$\Delta \log A^{\text{lump-sum}} \approx \underbrace{\Omega_d \mathbb{E}_\chi \left[\frac{\log s_h^0}{\theta - 1}\right]}_{\text{1st order}} \underbrace{- \frac{1}{2}\Omega_d^2 Var_\chi \left(\frac{\log s_h^0}{\theta - 1}\right) + \frac{1}{2}(\rho - 1)\Omega_d(1 - \Omega_d)\mathbb{E}_\chi \left[\left[\frac{\log s_h^0}{\theta - 1}\right]^2\right]}_{\text{2nd order with lump-sum taxation}},$$

this expression is identical to Equation (7) in Example 4 when there is no leisure, $\Omega_d = 1$. The first and second summands are the same as in (7) but are now scaled by Ω_d to account for the fact that households also consume leisure. The final summand, which is absent in (7), accounts for complementarities/substitutabilities between consumption and leisure. If consumption and leisure are complements, $\rho < 1$, then a negative shock to consumption caused by autarky reduces the value of leisure through complementarity.

Linear Taxation. Now consider the case where lump-sum taxation is unavailable so that lump-sum transfers must be non-negative: $\boldsymbol{T} \geq 0$, financed by a uniform consumption tax. Proposition 8 now implies that, to a second-order approximation,

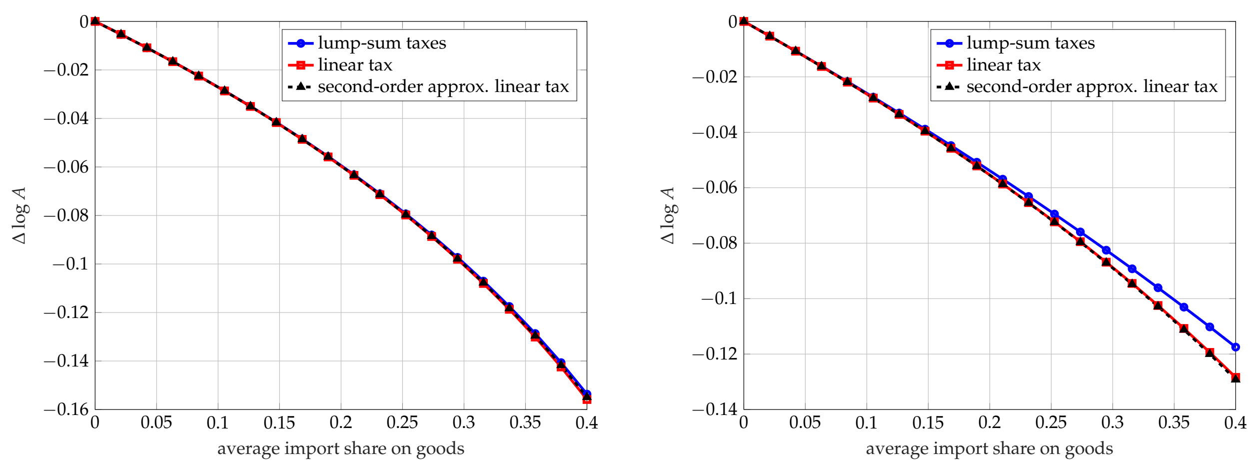
$$\Delta \log A^{\text{linear tax}} \approx \Delta \log A^{\text{lump-sum}} - \underbrace{\frac{1}{2}\rho\Omega_d(1 - \Omega_{hd})(d\log\tau^*)^2}_{\text{2nd order losses from distorting taxes}},$$

where τ^* is the optimal consumption tax in Equation (11).

Index the two households by h and h' and suppose that $s_{hd}^0 < s_{h'd}^0$. This means that, in the decentralized equilibrium, household h is more negatively affected by the trade shock than h'. In this case, the optimal feasible tax-and-transfer from (11) sends all collected tax revenues to h. Furthermore, to a first-order, the tax required for the compensation is $d \log \tau^* = \frac{\chi_h}{\theta - 1} \left[\log s^0_{hd} - \log s^0_{h'd} \right] > 0$.

The required consumption tax is larger the bigger is the heterogeneity in exposure to the trade shock, and the larger is household h's share of aggregate income. A given tax is more distorting the higher is ρ, which controls substitution between consumption and leisure (ρ can be interpreted as the Frisch elasticity of labor supply), and the closer is Ω_d to 1/2. If Ω_d is equal to either one (households do not value leisure) or zero (households do not value consumption), then there is no distortion from the tax.

Example 7 numerically illustrates the performance of the second-order approximation to the exact solution with distortionary linear taxes, and compares them both to the solution with lump-sum taxes. The second-order approximation performs well even for large shocks. Panel 6a uses $\rho = 0.5$, so consumption and leisure are complements and the Frisch elasticity of labor supply is a reasonable 0.5. Since ρ is low, distortionary taxes are able to achieve an outcome that is roughly as good as lump-sum taxes. Panel 6b uses a much higher $\rho = 3$. In this case, the gap between the lump-sum and linear taxation scenarios is larger since consumption taxes reduces labor and increase leisure, which causes efficiency to fall.



(a) $\rho = 0.5$

(b) $\rho = 3.0$

Figure 6: A numerical example of the losses from autarky with and without distortionary redistribution. The other parameter values are $\Omega_d = 0.5$, $\chi_h = 0.5$, $\theta_h = 3$, $s^0_{hd} = 3s^0_{h'd}$.

To summarize: losses from autarky are larger if some households are more badly affected than others, especially if efficient redistributive tools are not available to compensate the households that are more badly affected.

In the previous example, households are differentially affected by the trade shock because they have different preferences (e.g. some households consume more imports than others). In the next example, we instead consider a situation where households are differentially affected by a technology shock because they have different sources of income — the relative real wage between households changes in response to technological change.

Example 8 (Skill-Biased Technical Change). We now consider a simple example with skill-biased technical change that raises the real wage of high-skill workers but lowers the real-wage for low-skill workers. We compare how the response of aggregate efficiency changes depending on the redistributive tools available. Suppose that output (and consumption) are a CES aggregate of the output of manufacturing and services:

$$c = y = \left[\gamma_1^{\frac{1}{\rho}} y_m^{\frac{\rho-1}{\rho}} + (1-\gamma_1)^{\frac{1}{\rho}} y_s^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}},$$

where each sector's output is a CES aggregate of low- and high-skill labor

$$y_o = \left[\alpha_o^{\frac{1}{\sigma}} \left(z_{o1} l_{o1}\right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha_o)^{\frac{1}{\sigma}} \left(z_{o2} l_{o2}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}},$$

where l_{o1} is low- and l_{o2} is high-skill labor. The resource constraints are that

$$\sum_h c_h = c, \qquad \sum l_{o1} = l_1, \qquad \sum l_{o2} = l_2.$$

We assume that workers are much more substitutable than sectors: $\rho \ll \sigma$. We also assume that manufacturing is more intensive in low-skill labor use than services.

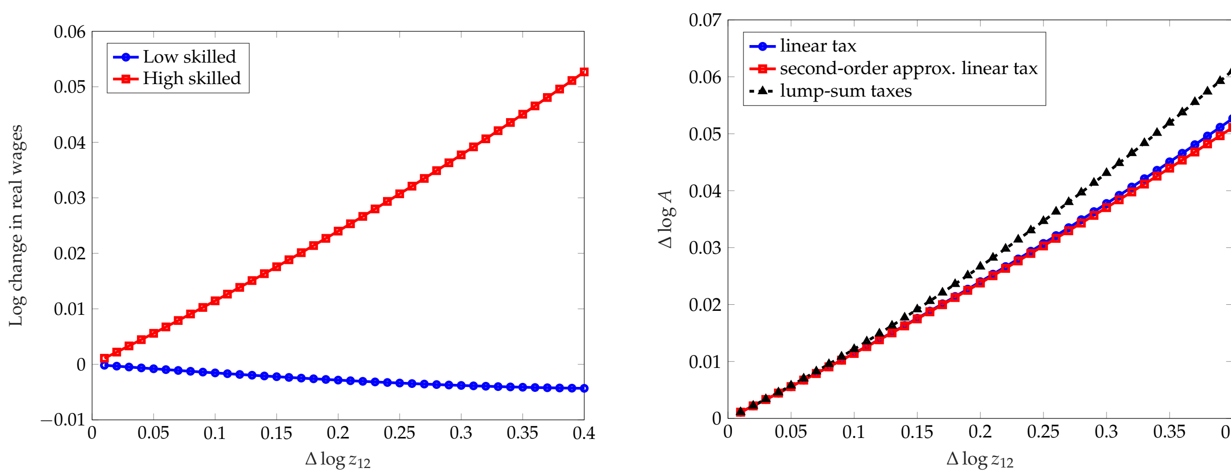
Consider an increase in automation or the productivity of capital, which we capture via an increase in the productivity of high-skill labor in manufacturing: $\Delta \log z_{m2} > 0$. This is a reduced-form representation for the idea that high-skill labor in manufacturing is equipped by capital, and hence an increase in the quality of capital makes high-skill more productive.[40]

Again, we contrast two scenarios: (1) lump-sum taxation is available, (2) lump-sum transfers must be non-negative and the government can only levy a linear tax on machine

[40]For example, high-skill labor and capital are combined in a Leontief nest together called equipped labor, and then equipped labor is substitutable with low-skill labor. We can then think of altering the productivity of equipped labor by varying the productivity of capital.

use in manufacturing, which we capture as a linear tax, τ, on manufacturing's use of high-skill labor.

Figure 7 illustrates the results in a numerical example. Panel 7a shows that skill-biased technical change raises the real wage for high-skill workers and lowers them for low-skill workers in the decentralized equilibrium. The fact that low-skill wages decline means that they need to be compensated via transfers financed by either lump-sum or distortionary taxes. Panel 7b shows the increase in efficiency depending on which taxes are used. As expected, the increase in aggregate efficiency is lower if only distortionary redstributive tools are available. Panel 7b also shows that the second-order approximation is very accurate. In the absence of any redistributive tools whatsoever, aggregate efficiency in this example actually declines because the low-skill workers are worst off and there is no feasible way to compensate them.



(a) real wages

(b) aggregate efficiency

Figure 7: A numerical example of skill-biased technical change. The parameter values are $\rho = 1$, $\sigma = 8$, $\gamma = 0.5$, $\alpha_{m1} = 0.9$, and $\alpha_{s1} = 0.5$. We normalize steady-state quantities so that the CES share parameters are equal to expenditure shares in the status-quo.

---

7 Explicit Characterization of Compensated Equilibrium

Theorem 2 and Theorem 3 show that calculating changes in aggregate efficiency can be boiled down to solving for the compensated equilibrium. This section provides some formulas for calculating variables in the compensated equilibrium. To do so, we rely on the results in Baqaee and Farhi (2020), which provide a characterization of solutions of representative agent economies with wedges.

For concreteness, assume that all production and utility functions are nested-CES.[41] To make the notation more compact, represent the economy in such a way that each producer, i, is associated with a single elasticity of substitution θ_i (by treating each sub-nest as a separate producer)

7.1 Input-Output Notation

Stack the expenditure shares of the representative household, all producers, and all factors into the $(H+N+F)\times(H+N+F)$ input-output matrix Ω. The first H rows correspond to the households consumption baskets. The next N rows correspond to the expenditure shares of each producer on every other producer and factor. The last F rows correspond to the expenditure shares of the primary factors (which are all zeros, since primary factors do not require any inputs). With some abuse of notation, the heterogeneous agent input-output matrix can be written as

$$\Omega = \left[\begin{array}{ccc|ccc|ccc}
0 & \cdots & 0 & b_{11} & \cdots & b_{1N} & 0 & \cdots & 0 \\
\vdots & \cdots & \vdots & & \cdots & & & \cdots & \\
0 & \cdots & 0 & b_{H1} & \cdots & b_{HN} & 0 & \cdots & 0 \\
\hline
0 & \cdots & 0 & \Omega_{11} & \cdots & \Omega_{1N} & \Omega_{1N+1} & \cdots & \Omega_{1N+F} \\
\vdots & \cdots & \vdots & & \ddots & & & & \\
0 & \cdots & 0 & \Omega_{N1} & & \Omega_{NN} & \Omega_{NN+1} & \cdots & \Omega_{NN+F} \\
\hline
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{array}\right]$$

The Leontief inverse matrix is the $(H+N+F)\times(H+N+F)$ matrix defined as

$$\Psi \equiv (I-\Omega)^{-1} = I + \Omega + \Omega^2 + \ldots,$$

where I is the identity matrix. The Leontief inverse matrix $\Psi \geq I$ records the *direct and indirect* exposures through the supply chains in the production network.

Denote the distribution of expenditures by each household by χ, which is an $(H+N+F)\times 1$ vector. The first H elements are equal to each household's share of aggregate consumption expenditures, and the remaining $N+F$ elements are all zeros. As a matter of ac-

[41] Non-CES economies can be analyzed in a similar way following the non-CES extensions in Baqaee and Farhi (2019c).

counting identities, the vector of Domar weights satisfies:

$$\lambda' = \chi' \Psi.$$

In this equation λ' is a $(H + N + F) \times 1$ vector. The first H elements are equal the expenditures of each household relative to aggregate consumption expenditures, χ', the next $N + F$ elements are equal to the sales of each good and factor relative to aggregate consumption expenditures.

Let μ and τ denote the diagonal matrices whose iith element is equal to μ_i and τ_i respectively. Define the cost-based Leontief inverse to be

$$\tilde{\Psi} = (I - (\tau\mu)^{-1}\Omega)^{-1}.$$

Note that the cost-based Leontief inverse coincides with Ψ in the absence of wedges. Intuitively, $\tilde{\Psi}$ is a version of the Leontief inverse that calculates exposures of i to j in terms of cost shares rather than revenue shares (revenues exceed costs if wedges and taxes are greater than one).

For any non-negative vector a, define

$$Cov_a(b, c) = \mathbb{E}_a[bc] - \mathbb{E}_a[b]\mathbb{E}_a[c] = \sum_i \frac{a_i}{\sum_{i'} a_{i'}} b_i c_i - \sum_i \frac{a_i}{\sum_{i'} a_{i'}} b_i \sum_i \frac{a_i}{\sum_{i'} a_{i'}} c_i,$$

where $\mathbb{E}_a[\cdot]$ denotes averages of vectors weighted by the elements of a. For any matrix X, denote its ith row and column by $X_{(i,:)}$ and $X_{(:,i)}$.

7.2 Differential Hat-Algebra

The next proposition characterizes compensated variables in terms of initial expenditure shares, wedges, and shocks.

Proposition 9 (Differential Equations for Compensated Equlibrium)**.** *Assuming interior outcomes, the compensated equilibrium satisfies the following system of differential equations. For each $i \in H + N + F$, the compensated price satisfies*

$$d \log p_i^{comp} = \sum_j \tilde{\Psi}_{ij}^{comp} [d \log \mu_j \tau_j^* - d \log z_j] + \sum_{f \in F} \tilde{\Psi}_{if}^{comp} d \log \lambda_f^{comp}. \tag{13}$$

Compensated Domar weights for goods and factors satisfy

$$d\lambda_l^{comp} = \sum_j \lambda_j^{comp}(1-\theta_j)\mu_j^{-1} Cov_{\Omega^{comp}_{(j,:)}}\left(d\log p^{comp}, \Psi^{comp}_{(:,l)}\right) + Cov_{\chi^{comp}}\left(d\log \chi^{comp}, \Psi^{comp}_{(:,l)}\right) - \sum_j \lambda_j \left(\Psi_{jl} - \mathbf{1}[j=l]\right) d\log \mu_j \tau_j^*. \quad (14)$$

Changes in compensated expenditure shares for household h satisfy

$$d\log \chi_h^{comp} = d\log p_h^{comp}, \quad (15)$$

where $d\log p_h^{comp}$ is the price of the consumption bundle for household h. The compensated input-output matrix satisfies

$$d\Omega_{ij}^{comp} = (1-\theta_i)\left(d\log p_j^{comp} - \mathbb{E}_{\Omega^{comp}_{(i,:)}}[d\log p^{comp}]\right) - d\log \mu_i. \quad (16)$$

Finally, $d\log y_i^{comp}$ is given by $d\log \lambda_i^{comp} - d\log p_i^{comp}$. The initial conditions are given by Lemma 1 that all prices and expenditures are equal to the ones in the competitive equilibrium for $t = 0$.

Equation (13), (14), and (16) are standard and identical to expressions in Baqaee and Farhi (2020). They are loglinearizations of marginal cost-functions, market clearing conditions, and demand curves respectively. The key equation, which distinguishes the compensated equilibrium from the decentralized equilibrium is (15). Whereas in the decentralized equilibrium changes in household expenditures are determined by changes in the income of each household, in the compensated equilibrium, they are determined by the choices of the Hicksian representative agent (who tries to equate homothetized utilities across agents). The term $d\log p_h^{\text{comp}}$, which is pinned down by (13), is the change in the compensated price index of household h.

The taxes $\boldsymbol{\tau}^*(t)$ are given by the maximizers of the problem in (12). If only lump-sum transfers are used for redistribution, as in Sections 4 and 5, then $\boldsymbol{\tau}^*(t) = 0$, and Proposition 9 fully characterizes the compensated equilibrium in terms of exogenous parameters: $\boldsymbol{z}(t)$ and $\boldsymbol{\mu}(T)$. If lump-sum transfers are unavailable, then solving for $\tau^*(t)$ requires specifying more details about the set of available tax instruments. Specifically, we would need to add the log-linearized first-order conditions for the tax instruments from (12) as additional equations in Proposition 9 to pin down how τ^* evolves.

There is one case where this optimization problem can be avoided. If there are only $H-1$ taxes available, and outcomes are interior, then (15) can pin down $\tau^*(t)$. For example, suppose that there are $H-1$ taxes, and the share of revenues from the ith tax sent to household h are given by α_{ih}:

$$T_h(t) = \sum_i \alpha_{ih} \left(1 - \frac{1}{\tau_i^*(t)}\right) \lambda_i(t).$$

Log-differentiating household h's budget constraint gives:

$$d\log \chi_h^{\text{comp}} = \sum_f \frac{\omega_{hf}}{\chi_h} d\log \lambda_f^{\text{comp}} + \frac{dT_h}{\chi_h^{\text{comp}}},$$

differentiating $T_h(t)$ above, and substituting it into the log-linearized budget constraint gives $H-1$ additional equations which, assuming regularity conditions, will pin down $d\log \boldsymbol{\tau}^*$.

Generally, solving the system of linear equations in Proposition 9 requires inverting a system of equations. When there is a single primary factor of production and we evaluate these derivatives at a perfectly competitive point, then the change in efficiency can be solved out easily up to a second-order.

Proposition 10 (Aggregate Efficiency with One Factor)**.** *Consider a competitive economy with a single primary factor of production. The change in aggregate efficiency in response to a vector of productivity shocks, $\Delta \log \boldsymbol{z}$ and changes in wedges $\Delta \log \boldsymbol{\mu}$ is*

$$\Delta \log A \approx \sum_i \lambda_i \Delta \log z_i + \frac{1}{2} \sum_{i \in N+H} \lambda_i (\theta_i - 1) Var_{\Omega(i,:)} \left(\sum_k \Psi_{(:,k)} \log z_k \right)$$
$$- \frac{1}{2} \sum_{i \in N+H} \lambda_i \theta_i Var_{\Omega(i,:)} \left(\sum_k \Psi_{(:,k)} \Delta \log(\mu_k \tau_k^*) \right).$$

to a second-order approximation in $\Delta \log z$ and $\Delta \log \mu$.

There are three summands. The first one is just Hulten's theorem. The second summand is a nonlinear adjustment due to changes in Domar weights. The second summand is also equal to: $1/2 \sum_j \partial \lambda_k^{\text{comp}} / \partial \log z_j \Delta \log z_j \Delta \log z_k$. If the compensated Domar weight for k rises due to productivity shocks, then the shock to k is more important. This happens if exposure to k is heterogeneous, captured by the variance term, and if elasticities of substitution, θ_i, are far from unity. The final summand are the Harbeger triangles caused by the taxes and wedges. The triangles are larger the higher are elasticities of substitution, θ_i, and the more heterogeneous are exposures to the taxes and wedges, captured by the variance terms.

8 Conclusion

This paper defines a measure of aggregate efficiency using the aggregate consumption-equivalent variation. We establish that this measure can be computed by solving for the equilibrium in a fictitious representative agent economy. This provides a method for translating theorems and tools about representative-agent economies to study aggregate efficiency in economies with heterogeneous agents. This includes Hulten (1978), Harberger (1964), Petrin and Levinsohn (2012), Arkolakis et al. (2012), and Baqaee and Farhi (2019c) and Baqaee and Farhi (2020).

In two stand-alone companion papers, we apply the theoretical results of this paper to contexts where household heterogeneity is of the utmost importance. Baqaee and Burstein (2025b) consider losses in aggregate efficiency from incomplete risk sharing within and across borders. Baqaee and Burstein (2025a) characterize aggregate efficiency in random utility models with discrete choice, focusing on spatial economies, where households make different choices due to differences in their preferences. An interesting extension, which we do not pursue in this paper but pursue in ongoing work, is studying policy problems where maximizing aggregate efficiency is the objective of the policymaker.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Alchian, A. A. and B. Klein (1973). On a correct measure of inflation. *Journal of Money, Credit and Banking 5*(1), 173–191.

Allais, M. (1979). La théorie générale des surplus.

Antras, P., A. de Gortari, and O. Itskhoki (2017). Globalization, inequality and welfare. *Journal of International Economics 108*(C), 387–412.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory 2*(3), 244–263.

Auer, R., A. Burstein, S. Lein, and J. Vogel (2024). Unequal expenditure switching: Evidence from switzerland. *Review of Economic Studies 91*(5), 2572–2603.

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

Baqaee, D. and A. Burstein (2025b). Efficiency costs of market incompleteness. Technical report.

Baqaee, D., A. Burstein, and Y. Koike-Mori (2024). Sufficient statistics for measuring forward-looking welfare. Technical report, National Bureau of Economic Research.

Baqaee, D. and E. Farhi (2019a, July). Networks, barriers, and trade. Working Paper 26108, National Bureau of Economic Research.

Baqaee, D. and E. Farhi (2019b). A short note on aggregating productivity. Technical report, National Bureau of Economic Research.

Baqaee, D. and E. Rubbo (2023). Micro propagation and macro aggregation. *Annual Review of Economics 15*(1), 91–123.

Baqaee, D. R. and A. Burstein (2023). Welfare and output with income effects and taste shocks. *The Quarterly Journal of Economics 138*(2), 769–834.

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

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

Baqaee, D. R. and E. Farhi (2024). Networks, barriers, and trade. *Econometrica 92*(2), 505–541.

Bergson, A. (1938, February). A reformulation of certain aspects of welfare economics. *The Quarterly Journal of Economics 52*(2), 310–334.

Bhandari, A., D. Evans, M. Golosov, and T. Sargent (2021). Efficiency, insurance, and redistribution effects of government policies. Technical report, Working paper.

Bigio, S. and J. La'O (2016). Financial frictions in production networks. Technical report.

Blackorby, C. and D. Donaldson (1990). A review article: The case against the use of the sum of compensating variations in cost-benefit analysis. *canadian Journal of Economics*, 471–494.

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

Bornstein, G. and A. Peter (2024). Nonlinear pricing and misallocation. Technical report, National Bureau of Economic Research.

Carvalho, V. M. and A. Tahbaz-Salehi (2019). Production networks: A primer. *Annual Review of Economics 11*, 635–663.

Comin, D., D. Lashkari, and M. Mestieri (2021). Structural change with long-run income and price effects. *Econometrica 89*(1), 311–374.

Cornes, R. (1992). *Duality and modern economics*. CUP Archive.

Dávila, E. and A. Schaab (2022). Welfare assessments with heterogeneous individuals.

Technical report, National Bureau of Economic Research.

Dávila, E. and A. Schaab (2023). Welfare accounting. Technical report, National Bureau of Economic Research.

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

Debreu, G. (1954). A classical tax-subsidy problem. *Econometrica: Journal of the Econometric Society*, 14–22.

Del Canto, F. N., J. R. Grigsby, E. Qian, and C. Walsh (2023). Are inflationary shocks regressive? a feasible set approach. Technical report, National Bureau of Economic Research.

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

Dupuit, A. J. É. J. (1844). De la mesure de lútilité des travaux publics. *Annales des ponts et chaussées 8*.

Fagereng, A., M. Gomez, E. Gouin-Bonenfant, M. Holm, B. Moll, and G. Natvik (2022). Asset-price redistribution. Technical report, Working Paper.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the royal statistical society series a: statistics in society 120*(3), 253–281.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Galle, S., A. Rodríguez-Clare, and M. Yi (2023). Slicing the pie: Quantifying the aggregate and distributional effects of trade. *The Review of Economic Studies 90*(1), 331–375.

Harberger, A. C. (1954). Monopoly and resource allocation. In *American Economic Association, Papers and Proceedings*, Volume 44, pp. 77–87.

Harberger, A. C. (1964). The measurement of waste. *The American Economic Review 54*(3), 58–76.

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

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

Heathcote, J. and H. Tsujiyama (2021). Optimal income taxation: Mirrlees meets ramsey. *Journal of Political Economy 129*(11), 3141–3184.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly journal of economics 135*(3), 1209–1318.

Hicks, J. R. (1939). The foundations of welfare economics. *The economic journal 49*(196), 696–712.

Hopenhayn, H. A. (2014, August). On the Measure of Distortions. NBER Working Papers 20404, National Bureau of Economic Research, Inc.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *The quarterly journal of economics 124*(4), 1403–1448.

Hulten, C. R. (1973). Divisia index numbers. *Econometrica: Journal of the Econometric Society*, 1017–1025.

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

Johansson, P.-O., B. Kriström, et al. (2022). The boadway paradox revisited: The case of compensated equilibrium. *Center for Environmental and Resource Economics, CERE, The Boadway Paradox Revisited: The Case of Compensated Equilibrium (June 13, 2022)*.

Jones, C. (2002). The" boadway paradox" revisited.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The economic journal 49*(195), 549–552.

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

Liu, E. (2017). Industrial policies and economic development. Technical report.

Lucas, R. E. (1987). *Models of business cycles*, Volume 26. Basil Blackwell Oxford.

Luenberger, D. G. (1996). Welfare from a benefit viewpoint. *Economic Theory 7*(3), 445–462.

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

Petrin, A. and J. Levinsohn (2012). Measuring aggregate productivity growth using plant-level data. *The RAND Journal of Economics 43*(4), 705–725.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics 11*(4), 707–720.

Rodríguez-Clare, A., M. Ulate, and J. P. Vasquez (2022). *Trade with nominal rigidities: Understanding the unemployment and welfare effects of the China shock*. CESifo, Center for Economic Studies & Ifo Institute.

Samuelson, P. A. (1983). *Foundations of economic analysis*, Volume 197. Harvard University Press Cambridge, MA.

Schulz, K., A. Tsyvinski, and N. Werquin (2023). Generalized compensation principle. *Theoretical Economics 18*(4), 1665–1710.

Scitovszky, T. (1941). A note on welfare propositions in economics. *The Review of Economic Studies 9*(1), 77–88.

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

Appendix A Representative Agent

We follow the definition of a positive representative agent in Mas-Colell et al. (1995). We say that u^{RA} is a representative agent if the Marshallian demand curves of the representative agent, given prices and total income, coincide with equilibrium allocations given the same prices and aggregate income:

$$\arg\max_{\boldsymbol{c}}\{u^{RA}(\boldsymbol{c}) : \sum_i p_i(t)c_i \leq I(t)\} = \sum_h \arg\max_{\boldsymbol{c}_h}\{u_h(\boldsymbol{c}_h) : \sum_i p_i(t)c_{hi} \leq I_h(t)\}.$$

Appendix B Expenditure Function of Hicksian RA

The following proposition characterizes the expenditure function of the Hicksian representative agent.

Proposition 11 (Dual Representation of Hicksian Representative Agent)**.** *The expenditure function associated with $U(\boldsymbol{c})$, in Definition 2, denoted by $E(p, U)$ is*

$$E(\boldsymbol{p}, U) = \left(\sum_{h\in H} e_h(\boldsymbol{p}; u_h(\boldsymbol{c}_h^0))\right) U.$$

By Shephard's lemma, the budget share of the Hicksian representative agent on good i, denoted b_i^{comp}, is

$$b_i^{comp}(\boldsymbol{p}) = \frac{\partial \log E(\boldsymbol{p}, U)}{\partial \log p_i} = \sum_h \frac{e_h(\boldsymbol{p}, u_h^0)}{\sum_{h'} e_{h'}(\boldsymbol{p}, u_{h'}^0)} b_{hi}(\boldsymbol{p}, u_h^0),$$

where $b_{hi}(\boldsymbol{p}, u_h^0)$ is the compensated budget share of household i at the status-quo indifference curve $u_h^0 \equiv u_h(\boldsymbol{c}_h^0)$.

In words, the Hicksian representative agent's spending on each good i is the average compensated budget share of all households, where each household is weighted according to its compensating variation, $e_h(\boldsymbol{p}, u_h^0)$.

Given compensated aggregate budget shares, $b_i^{\text{comp}}(\boldsymbol{p})$, we can solve for equilibrium variables in the compensated equilibrium including prices $\boldsymbol{p}^{\text{comp}}$. Setting aggregate spending to be the numeraire in the compensated equilibrium, and using Theorem 2, we know that

$$A(t) = U(t) = \frac{1}{\left(\sum_{h\in H} e_h(\boldsymbol{p}^{\text{comp}}; u_h(\boldsymbol{c}_h^0))\right)}.$$

Appendix C Proofs

Proof of Theorem 1. Denote the solution to (1) by ϕ^*, and an allocation that attains this solution (it does not need to be unique) by $\boldsymbol{c}^* \in \exp(-\phi^*)\mathcal{C}$. By the definition of ϕ^* and $U(\cdot)$, and given local nonsatiaton, $U(\boldsymbol{c}^*) = U(\boldsymbol{c}^0) = 1$. Denote the solution to $V(\mathcal{C})$ by $\boldsymbol{c}^{**}$, with $\boldsymbol{c}^{**} \in \mathcal{C}$. Since $U(\cdot)$ is homogeneous of degree 1, it follows that $V(\exp(-\phi^*))\mathcal{C}) = \exp(-\phi^*)V(\mathcal{C})$, with solution $\exp(-\phi^*)\boldsymbol{c}^{**}$. Note that $V\left(\exp(-\phi^*)\mathcal{C}\right) \geq 1$ because $\boldsymbol{c}^* \subseteq \exp(-\phi^*)\mathcal{C}$ and $U(\boldsymbol{c}^*) = 1$. Moreover, $V\left(\exp(-\phi^*)\mathcal{C}\right) = 1$ because, if it were strictly higher than 1, the solution to (1) would be higher than ϕ^*. It thus follows that:

$$V(\mathcal{C}) = \frac{V(\mathcal{C})}{V\left(\exp(-\phi^*)\mathcal{C}\right)} = \exp\left(\phi^*\right).$$

□

Proposition 12 (Equivalence of Kaldor-Hicks and Aggregate Efficiency)**.** *If there is one primary factor, so that relative prices are independent of demand, then* $A(t) = A^{KH}(t)$*.*

Other proofs to be added.