# Supplemental Appendix

## Nonlinear Pricing and Misallocation

Gideon Bornstein* Alessandra Peter†

June, 2025

# A Continuum of types

In this appendix, we set up the baseline model from Section 2 for an environment in which consumer tastes are drawn from a continuous distribution. We show that the propositions and proofs remain the same. We compare the quantitative results to the baseline calibration from Section 4. The implied price dispersion is somewhat smaller, but the allocation of goods closely resembles the two types model. In the last part, we restrict firms to offering 2 bundles only. We show that with CED preferences, the allocation of consumers to the two bundles is independent of firm productivity and quantify the model.

## A.1 Theory: Model setup

Household preferences are as before, with the only difference that taste shifter $\tau_{ij}$ are drawn from a cumulative distribution function $G(\tau)$ with support on $[1, \bar{\tau}]$. The CDF $G$ is continuously differentiable, and has non-decreasing hazard rate, $h(\tau) \equiv \frac{g(\tau)}{1-G(\tau)}$.[54]

**Firms.** Each firm $j$ chooses a pricing schedule $p(q)$ that maximizes expected profits. This pricing schedule also implies a mapping of consumer taste $\tau$ to a quantity purchased $q(\tau)$. Since firms cannot condition on type, they must ensure that consumers self-select into their type's bundle.

$$\max_{\{q_j(\tau), p_j(q)\}} \quad \int_\tau q_j(\tau)\left(p_j(q_j(\tau)) - c_j\right) dG(\tau) \tag{A.1}$$

$$q_j(\tau) \in \underset{q\geq 0}{\operatorname{argmax}} \left[\tau u(q) - \frac{p_j(q)q}{P}\right], \qquad \forall \tau$$

The set of constraints in Problem (A.1) states that each consumer type $\tau$ must prefer their allocation to not buying the good ($q = 0$, the IR constraint) and to buying any other positive quantity (the

*The Wharton School, University of Pennsylvania and NBER; gideonbo@wharton.upenn.edu
†New York University, CEPR, and NBER; alessandra.peter@nyu.edu

[54] This assumption is common and necessary in order to use the standard mechanism design tools, see Myerson (1981)

set of IC constraints).[55] We solve the problem of the firm using standard tools from the mechanism design literature (see Fudenberg and Tirole (1991)). In the solution to this problem, the individual rationality constraint binds for the lowest types ($\tau_{ij} = 1$), while the set of incentive compatibility constraints for these consumers are slack. For all other consumers, the only binding constraint is the downward local incentive compatibility constraint.

**Firm-level optimal prices and quantities.** The quantity sold to consumers of a particular taste $\tau$ is implicitly given by

$$\tau u'(q_j(\tau)) = \frac{c_j}{P} \frac{\tau}{\tau - [h(\tau)]^{-1}} \tag{A.2}$$

Firms choose a quantity $q_j(\tau)$ that equates the marginal utility of each consumer, $\tau u'(q_j(\tau))$, to the *effective cost* of the good. The effective cost consists of two components. First, the real marginal cost of producing the good is $c_j/P$. Second, selling an additional unit entails a *shadow* cost. In order to ensure that consumers with higher taste are still willing to purchase their designated quantity, the prices these consumers pay must go down.

In choosing the optimal quantity offered to consumers with taste $\tau$, the firm takes into account the measure of consumers with that given taste, $g(\tau)$, who will now purchase an additional unit, relative to the measure of consumers with a higher taste for the good, $1 - G(\tau)$, who must now be charged a marginally lower price. This is the hazard rate $h(\tau)$. The higher is the hazard rate, the higher is the measure of consumers with taste $\tau$ relative to consumers with higher tastes, and the lower is the shadow cost of selling an additional unit to consumers with taste $\tau$.

Markups charged by the firm are given by

$$\mu_{ij} = \psi\left(q_{ij}\right) \frac{\tau_{ij}}{\tau_{ij} - h^{-1}(\tau_{ij})} \left[1 - \frac{\int_0^i \tau_{kj} u'(q_{kj}) dk}{\tau_{ij} u(q_{ij})}\right] \tag{A.3}$$

The term $\psi(q)$ is the *social markup*, a term coined by Dhingra and Morrow (2019). If firms could perfectly price discriminate, they would extract the full consumer surplus from each of their consumers. The markup charged from each consumer would be equal to the social markup $\psi(q_{ij})$. With nonlinear pricing, firms are able to extract the full consumer surplus only of the consumers with the lowest taste. Consumers with a high taste on the other hand have a positive consumer surplus, which is necessary to achieve separation.

**Efficient allocation.** The first-best allocation solves the planner's problem as in Equation (2.6). The optimal allocations are given by

$$u'(q_{ij}^{\text{FB}}) = \frac{c_j}{\tau_{ij}} \frac{1}{P^{\text{FB}}}, \tag{A.4}$$

[55] As before, we assume that the distribution of tastes $G(\tau)$, the distribution of firm productivities $F(c)$ and preference parameters are such that all firms optimally choose to serve all types of consumers.

where $P^{\mathrm{FB}}$ is the inverse Lagrange multiplier on the aggregate resource constraint.

## A.2 Theory: Propositions and proofs

PROPOSITION **13.** *In equilibrium, there is a cut-off taste $\hat{\tau}$ for each good $j$ such that all consumers with $\tau > \hat{\tau}$ are allocated too much, and all consumers with $\tau < \hat{\tau}$ are allocated too little of the good.*

PROOF OF PROPOSITION **13.** From equations (A.2) and (A.4) we have that:

$$\frac{u'\left(q_{\tau j}\right)}{u'\left(q_{\tau j}^{FB}\right)} = \frac{P^{FB}}{P}\omega(\tau) \tag{A.5}$$

where $\omega(\tau) \equiv \frac{\tau}{\tau - [h(\tau)]^{-1}}$ Given that the hazard rate is non-decreasing, $\omega(\tau)$ in decreasing in $\tau$. Further, $\omega(\bar{\tau}) = 1$ and hence $\omega(\tau) \geq 1 \, \forall \tau$.

As in the model with two types, one of three cases must hold: (i) $P^{FB}/P > 1$ and therefore $q_{\tau j} < q_{\tau j}^{\mathrm{FB}} \; \forall \{\tau, j\}$, (ii) $P^{FB}/P \leq \omega(1)$ and therefore $q_{\tau j} \geq q_{\tau j}^{\mathrm{FB}} \; \forall \{\tau, j\}$, or (iii) $P^{FB}/P \in (\omega(1), 1)$ and therefore, for each $j$, $q_{\tau j} > q_{\tau j}^{\mathrm{FB}}$ for some $\tau$ and $q_{\tau j} < q_{\tau j}^{\mathrm{FB}}$ for others.

Only (iii) is consistent with labor market clearing. Let $\hat{\tau}$ be given by $\omega(\hat{\tau}) = P^{FB}$. Given we are in case (iii), $\hat{\tau} \in (1, \bar{\tau})$. It follows that first, for all $j$ $q_{\hat{\tau} j} = q_{\hat{\tau} j}^{FB}$. Second, since $\omega'(\tau) \leq 0$, $q_{\hat{\tau} j} > q_{\hat{\tau} j}^{FB}$ $\forall \, \tau > \hat{\tau}$ and $q_{\hat{\tau} j} < q_{\hat{\tau} j}^{FB} \quad \forall \, \tau < \hat{\tau}$.

■

PROPOSITION **14.** *Suppose preferences satisfy Assumption 1. Then, the equilibrium levels of firm-level production and employment are identical to the efficient allocation.*

PROOF OF PROPOSITION **14.**

From equation (A.2), it follows again that there is a unique level of the aggregate price index such that the labor market clears.

Let $\widetilde{P_j}$ be the aggregate price index such that the firm-level production of a firm with marginal cost $c_j$ in equilibrium is identical to its overall production in the efficient allocation.

$$\int_1^{\hat{\tau}} \left[q_j^{FB}(\tau) - q_j(\tau, \widetilde{P_j})\right] dG(\tau) - \int_{\hat{\tau}}^{\bar{\tau}} \left[q_j(\tau, \widetilde{P_j}) - q_j^{FB}(\tau)\right] dG(\tau) = 0 \tag{A.6}$$

By the same argument as in the Proof of Proposition 7, Assumption 1 implies that $\widetilde{P_j}$ is independent of firm cost hence total production is equal to first-best for all firms.

■

PROPOSITION **15.** *Suppose preferences satisfy Assumption 1. Then, the optimal firm-level subsidies and taxes are zero.*

Proof of Proposition 15.

Let's first set up the planner's problem using the primal approach. The planner chooses taxes and subsidies to all firms, $\{t_j\}$, such that its budget is balanced. By choosing taxes and subsidies the planner has control over the firm-level employment of all firms in the economy. We take the primal approach and write the planner's problem as follows:

$$\begin{aligned}
\max_{\{l_j, q_j(\tau)\}_{j=0}^1} \quad & \int_0^1 \int_\tau \tau u\left(q_j(\tau)\right) g(\tau) d\tau dj, && \text{(A.7)}\\
\text{s.t.} \quad & \frac{\tau}{\omega(\tau)} u'(q_j(\tau)) = \bar{\tau} u'(q_j(\bar{\tau})) && \forall\,(\tau, j)\\
& \int_\tau q_j(\tau) g(\tau) d\tau = \frac{l_j}{c_j}, && \forall\, j\\
& \int_0^1 l_j dj = 1.
\end{aligned}$$

Taking first order conditions, we obtain

$$[q_j(\tau)]: \qquad \tau u'\left(q_j(\tau)\right) g(\tau) - \mu_j(\tau) \frac{\tau u''(q_j(\tau))}{\omega(\tau)} g(\tau) = \theta_j g(\tau), \tag{A.8}$$

$$[q_j(\bar{\tau})]: \qquad \bar{\tau} u'\left(q_j(\bar{\tau})\right) g(\bar{\tau}) + \int_\tau \mu_j(\tau)\, \bar{\tau}\, u''(q_j(\bar{\tau}))\, g(\tau)\, d\tau = \theta_j g(\bar{\tau}), \tag{A.9}$$

$$[l_j]: \qquad \frac{\theta_j}{c_j} = \lambda, \tag{A.10}$$

where $\mu_j(\tau)$, $\theta_j$, and $\lambda$ are the (sets of)s Lagrange multipliers on the three constraints, respectively. Combining conditions (A.8) and (A.9), we get

$$\bar{\tau} u'\left(q_j(\bar{\tau})\right) g(\bar{\tau}) + \bar{\tau} \int_\tau \omega(\tau) u'(q_j(\tau)) \frac{u''(q_j(\bar{\tau}))}{u''(q_j(\tau))} g(\tau) d\tau = \left[ g(\bar{\tau}) + \bar{\tau} \int_\tau \frac{\omega(\tau)}{\tau} \frac{u''(q_j(\bar{\tau}))}{u''(q_j(\tau))} g(\tau) d\tau \right] \theta_j \tag{A.11}$$

Substituting out $\theta_j$ using (A.10) and using the fact that, under Assumption 1 $u''\left(q_j(\tau)\right)/u''\left(q_j\bar{\tau}\right) = \left(u'\left(q_j(\tau)\right)/u'\left(q_j(\bar{\tau})\right)\right)^{1+\eta}$, it follows that the optimality condition of the planner (A.11) holds at the market allocations characterized by (A.2). The resulting Lagrange multiplier $\lambda$ on the aggregate resource constraint is given by

$$\lambda = \frac{1}{P} \frac{g(\bar{\tau}) + \bar{\tau}^{-\eta} \int_\tau \omega(\tau)^{1-\eta} \tau^\eta g(\tau) d\tau}{g(\bar{\tau}) + \bar{\tau}^{-\eta} \int_\tau \omega(\tau)^{-\eta} \tau^\eta g(\tau) d\tau} \tag{A.12}$$

which is indeed independent of firm $j$. We conclude that the equilibrium allocations coincide with the constrained efficient allocation. Therefore, the optimal firm-level taxes and subsidies are all zero.

■

## A.3 Two bundles: Theory

Household preferences and production technology are as before. However, firms may only offer two bundles $(q_1, p_1)$ and $(q_\tau, p_\tau)$. Let $\hat{\tau}_j$ be the threshold below which consumers buy $q_1$ and above which $q_\tau$.[56]

### A.3.1 Market allocation

The firm's problem is given by

$$\begin{aligned}
\max_{\{q_{1j}, q_{\tau j}, p_{1j}, p_{\tau j}, \hat{\tau}_j\}} \quad & (p_{1j} - c_j) q_{1j} G(\hat{\tau}) + (p_{\tau j} - c_j) q_{\tau j} (1 - G(\hat{\tau})) \\
\text{s.t.} \quad & u(q_{1j}) = \frac{p_{1j} q_{1j}}{P}, \\
& \hat{\tau}_j u(q_{\tau j}) - \frac{p_{\tau j} q_{\tau j}}{P} = \hat{\tau}_j u(q_{1j}) - \frac{p_{1j} q_{1j}}{P}.
\end{aligned}$$

which uses the usual result that the IR constraint only binds for the lowest type and the IC only for the threshold consumer $\hat{\tau}$.

The optimal quantities $q_{1j}$ and $q_{\tau j}$ solve

$$\hat{\tau}_j u'(q_{\tau j}) = \frac{c_j}{P} \tag{A.13}$$

$$u'(q_{1j}) = \frac{G(\hat{\tau}_j)}{1 - \hat{\tau}_j (1 - G(\hat{\tau}_j))} \frac{c_j}{P} \tag{A.14}$$

Similar to the baseline model with two types, conditional on the aggregate price index P, the threshold type $\hat{\tau}_j$ is sold the optimal quantity and there is a wedge $\frac{G(\hat{\tau}_j)}{1-\hat{\tau}_j(1-G(\hat{\tau}_j))} > 1$ that distorts the allocation the lowest type downwards.

The threshold type who is indifferent between the two bundles solves:

$$\frac{c_j}{P} = \left( \hat{\tau}_j - \frac{1 - G(\hat{\tau}_j)}{g(\hat{\tau}_j)} \right) \frac{u(q_{\tau j}) - u(q_{1j})}{q_{\tau j} - q_{1j}} \tag{A.15}$$

The threshold type is a function of the hazard ratio associated with the distribution of consumer tastes $G(.)$. In general, it depends on the productivity of the firm. However, we show that, as long as preferences feature CED, the threshold type in independent of firm productivity. While all firms might choose to offer bundles that induce an inefficient allocation of consumers to quantities purchased, this distortion is constant across firms.

Proposition **16.** *Suppose preferences satisfy Assumption 1. Then the allocation of consumer tastes to the low and high bundles in the market equilibrium is identical for all firms. That is,* $\hat{\tau}_j = \hat{\tau}\ \forall j$.

[56] We maintain the assumption that firms choose to serve all customers. Note that in theory, a firm may choose its bundles in a way that excludes some low-taste consumers. We confirm in our quantitative analysis that no firm do not have incentive to exclude any customer.

### A.3.2 Efficient allocation

Suppose the social planner can also only choose two quantities for each firm, $q_{1j}^{FB}$ and $q_{\tau j}^{FB}$. The two quantities imply a cut-off type $\hat{\tau}_j^{FB}$. They are the solution to

$$\begin{aligned}
\max_{\{q_{1j}, q_{\tau j}, \hat{\tau}_j\}} \quad & \int_j \left[ \int_1^{\hat{\tau}_j} \tau u(q_{1j}) dG(\tau) + \int_{\hat{\tau}_j}^{\infty} \tau u(q_{\tau j}) dG(\tau) \right] dj \\
\text{s.t.} \quad & \int_j c_j \left( q_{1j} G(\hat{\tau}_j) + q_{\tau j}(1 - G(\hat{\tau}_j)) \right) = 1
\end{aligned}$$

As before, let $P^{FB}$ be the inverse Lagrange multiplier on the aggregate resource constraint. The optimal allocations and the cut-off types solve

$$\mathbb{E}\left[\tau | \tau \leq \hat{\tau}_j^{FB}\right] u'(q_{1j}^{FB}) = \frac{c_j}{P^{FB}} \tag{A.16}$$

$$\mathbb{E}\left[\tau | \tau \geq \hat{\tau}_j^{FB}\right] u'(q_{\tau j}^{FB}) = \frac{c_j}{P^{FB}} \tag{A.17}$$

$$\hat{\tau}_j^{FB} \frac{\left[u(q_{\tau j}^{FB}) - u(q_{1j}^{FB})\right]}{q_{\tau j}^{FB} - q_{1j}^{FB}} = \frac{c_j}{P^{FB}} \tag{A.18}$$

Conditional on a cut-off type $\hat{\tau}_j^{FB}$, the planner chooses the quantities that equate *expected* marginal utility to marginal cost for the set of households who purchase that bundle. The optimality condition for the cut-off type $\hat{\tau}_j^{FB}$ is similar to the market allocation with the exception that only the taste shifter enters.

From (A.18) it follows that, under CED, the cut-off type is the same for all firms also in the market allocation.

PROPOSITION **17.** *Suppose preferences satisfy Assumption 1. Then the allocation of consumer tastes to the low and high bundles in the first-best is identical for all firms. That is,* $\hat{\tau}_j^{FB} = \hat{\tau}^{FB}\ \forall j$.

In this environment, there are two dimensions of misallocation across firms. Relative to the social planner, the market allocation induces a different cut-off type. The set of consumers that purchase the high vs low-taste bundle are different. In addition, the two quantities offered are different. Consider for example the large bundle. The social planner chooses it to maximizes the average utility—net of costs—of households purchasing that bundles. The firm chooses it to maximize utility of the cut-off type.

Importantly, however, our main result of no misallocation across firms remains. Both types of misallocation across consumers do not depend on firm productivity. As long as preferences are CED, total production of each firm in the market allocation is identical to first-best.

PROPOSITION **18.** *Suppose preferences satisfy Assumption 1. Then, the equilibrium levels of firm-level production and employment are identical to the efficient allocation.*

## A.4 Two bundles: Proofs

Proof of Proposition 16.

Using Lemma 2, we can write the differences in utility and quantities as

$$q_{\tau j} - q_{1j} = \beta_1 \left(c_j/P\right)^{-\eta} \left[\hat{\tau}_j^{\eta} - \tilde{\tau}_j^{\eta}\right] \tag{A.19}$$

$$u(q_{\tau j}) - u(q_{1j}) = \frac{\eta}{\eta-1} \beta_1^{\frac{1}{\eta}} \left(c_j/P\right)^{1-\eta} \left[\hat{\tau}_j^{\eta-1} - \tilde{\tau}_j^{\eta-1}\right] \tag{A.20}$$

$$\tag{A.21}$$

where $\tilde{\tau}_j \equiv \frac{1-\hat{\tau}_j(1-G(\hat{\tau}_j))}{G(\hat{\tau}_j)}$. Taking ratios

$$\frac{u(q_{\tau j}) - u(q_j)}{q_{\tau j} - q_{1j}} = \frac{c_j}{P} \frac{\frac{\eta}{\eta-1}\beta_1^{\frac{1}{\eta}} \left[\hat{\tau}_j^{\eta-1} - \tilde{\tau}_j^{\eta-1}\right]}{\beta_1 \left[\hat{\tau}_j^{\eta} - \tilde{\tau}_j^{\eta}\right]} \tag{A.22}$$

Plugging into (A.15), the optimality condition for $\hat{\tau}_j$

$$\hat{\tau}_j - \frac{1-G(\hat{\tau}_j)}{g(\hat{\tau}_j)} = \frac{\beta_1 \left[\hat{\tau}_j^{\eta} - \tilde{\tau}_j^{\eta}\right]}{\frac{\eta}{\eta-1}\beta_1^{\frac{1}{\eta}} \left[\hat{\tau}_j^{\eta-1} - \tilde{\tau}_j^{\eta-1}\right]} \tag{A.23}$$

which is independent of $c_j$ and hence $\hat{\tau}_j = \hat{\tau}\ \forall j$

■

Proof of Proposition 17. Using Lemma (2) and the optimality conditions (A.16) and (A.17), rewrite (A.18) as

$$\frac{c_j}{P^{FB}} \frac{1}{\hat{\tau}_j^{FB}} = \frac{c_j}{P^{FB}} \frac{\frac{\eta}{\eta-1}\beta_1^{\frac{1}{\eta}} \left[\mathbb{E}\left[\tau | \tau \geq \hat{\tau}_j^{FB}\right]^{\eta-1} - \mathbb{E}\left[\tau | \tau \leq \hat{\tau}_j^{FB}\right]^{\eta-1}\right]}{\beta_1 \left[\mathbb{E}\left[\tau | \tau \geq \hat{\tau}_j^{FB}\right]^{\eta} - \mathbb{E}\left[\tau | \tau \leq \hat{\tau}_j^{FB}\right]^{\eta}\right]} \tag{A.24}$$

As before, the $\frac{c_j}{P^{FB}}$ cancel and the resulting cut-off type $\hat{\tau}_j^{FB}$ is constant across firms.

■

Proof of Proposition 18.

Let $\widetilde{P}_j$ be the aggregate price index such that the firm-level production of a firm with marginal cost $c_j$ in equilibrium is identical to its overall production in the efficient allocation:

$$G(\hat{\tau})\, q_{1j}(\widetilde{P}_j) + (1-G(\hat{\tau}))\, q_{\tau j}(\widetilde{P}_j) = G(\hat{\tau}^{FB})\, q_{1j}^{FB} + (1-G(\hat{\tau}^{FB}))\, q_{\tau j}^{FB} \tag{A.25}$$

$$G(\hat{\tau}) \left[q_{1j}(\widetilde{P}_j) - q_{1j}^{FB}\right] + (1-G(\hat{\tau})) \left[q_{\tau j}(\widetilde{P}_j) - q_{\tau j}^{FB}\right] = \left[G(\hat{\tau}) - G(\hat{\tau}^{FB})\right] \left[q_{\tau j}^{FB} - q_{1j}^{FB}\right] \tag{A.26}$$

Note that here we already used the fact that $\hat{\tau}$ and $\hat{\tau}^{FB}$ are both independent of $c_j$.

For the remainder of the proof, we follow the exact same steps as in the proof of Proposition 4. The only difference in the expressions for excess labor is the last term, $\left[G(\hat{\tau}) - G(\hat{\tau}^{FB})\right] \left[q_{\tau j}^{FB} - q_{1j}^{FB}\right]$.

Table A.1: Calibrated moments and parameters when firms offer 2 bundles

| Parameter | | Model | | Moment | | Model | |
|---|---|---|---|---|---|---|---|
| | | Benchm. | 2 sizes | | Data | Benchm. | 2 sizes |
| $\bar{\tau}$ | Highest taste | 1.19 | 1.22 | Package size dispersion | 0.44 | 0.44 | 0.44 |
| $\theta$ | Pareto shape | 3.14 | 3.12 | Sales share top 5% | 0.73 | 0.73 | 0.73 |
| $\eta$ | Elasticity of differences | 4.15 | 4.11 | Aggregate markup | 1.3 | 1.3 | 1.3 |
| $\beta_0$ | Departure from CES | 0.02 | 0.02 | Markup elast. w.r.t. size | 0.03 | 0.03 | 0.03 |

Under CED, this term is proportional to the firm's cost, with the same factor of proportionality as the difference between market and first-best, $\left[q_{\tau j}(\widetilde{P}_j) - q_{\tau j}^{FB}\right]$. Hence, $\widetilde{P_j}$ equates total labor demand of any firm to the first-best.

■

## A.5 Two bundles: Quantification

The model moments and calibrated parameters are displayed in Table A.1.

The misallocation costs are summarized in Table A.2. The welfare costs from misallocation with two package sizes are similar to our benchmark results: 0.71% relative to 0.68%.

Table A.2: Welfare Costs of Misallocation

| Baseline Model | 2 sizes |
|---|---|
| 0.68% | 0.71% |

**Notes:** This table reports the welfare costs in the decentralized equilibrium relative to the efficient allocation for our benchmark model (column 1), and the model in which firms are restricted to choosing 2 sizes only. All welfare costs are measured in consumption equivalent terms—that is, the uniform decline in consumption that would make households indifferent between the two equilibria.

# B Kimball preferences

In this section, we describe the model and its equilibrium with Kimball preferences. We map the Kimball aggregator to the canonical second-degree price discrimination problem with continuum of types.

## B.1 Canonical second-degree price discrimination results

There is a continuum of types $\tau \in [1, \bar{\tau}]$ which are drawn according to a pdf $g(\tau)$. We solve for the problem of a firm facing a constant marginal cost which is equal to $c$. Let the utility of consumer with type $\tau$ be given by

$$U(\tau) = b(q, \tau) - T,$$

where $q$ is the amount consumed and $T$ is the amount paid to the firm. We assume that the cross derivative is positive, $b_{q\tau}(\cdot) > 0$.

Following the standard derivations, we obtain the following:[57]

$$\frac{\partial b(q, \tau)}{\partial q} = c + h(\tau)^{-1} \frac{\partial^2 b(q, \tau)}{\partial \tau \partial q} \tag{B.1}$$

where $h(\tau) \equiv \frac{g(\tau)}{1 - G(\tau)}$. And

$$T(\tau) = b(q(\tau), \tau) - \int_1^\tau \frac{\partial b(q(t), t)}{\partial t} dt \tag{B.2}$$

## B.2 Mapping Kimball preferences to the canonical problem

With Kimball preferences, the aggregate consumption is implicitly defined as follows

$$\int \tau_{ij} \Upsilon \left( \frac{q_{ij}}{Q_i} \right) dj = 1, \tag{B.3}$$

Instead of $j$ notation, let's work directly with productivity of the firm ($c$). Then, the equation becomes

$$\int \int \tau \Upsilon \left( \frac{q(c, \tau)}{Q_i} \right) dG(\tau) dF(c) = 1 \tag{B.4}$$

This shows that $Q_i$ doesn't vary by $i$, where we used the fact that taste draws are iid there is a continuum of firms. So we will use $Q$ from here on:

$$\int \int \tau \Upsilon \left( \frac{q(c, \tau)}{Q} \right) dG(\tau) dF(c) = 1 \tag{B.5}$$

We want to map this into the general formulation of section B.1. We shall use the following Lemma.

[57] For detailed derivations, see, for example, https://faculty.haas.berkeley.edu/hermalin/continuous_2nd_degree.pdf.

Lemma **3.** *The utility from consuming $q$ instead of 0 for a consumer with taste $\tau$ is*

$$\tau Q D\left[\Upsilon\left(\frac{q}{Q}\right)-\Upsilon(0)\right]. \tag{B.6}$$

*where*

$$D=\frac{1}{\int\int\tau\Upsilon'\left(\frac{q(c,\tau)}{Q}\right)\frac{q(c,\tau)}{Q}dG(\tau)dF(c)}. \tag{B.7}$$

Proof. Aggregate consumption is defined by

$$\int\tau_{ij}\Upsilon\left(\frac{q_{ij}}{Q_i}\right)dj=1, \tag{B.8}$$

We start by deriving an expression for $\frac{\partial Q}{\partial q_{ij}}$.Totally differentiating, we get

$$\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz\frac{1}{Q^2}dQ=\frac{1}{Q}\Upsilon'\left(\frac{q_{ij}}{Q}\right)\tau_{ij}dq_i, \tag{B.9}$$

which yields

$$\frac{\partial Q}{\partial q_{ij}}=Q\frac{\tau_{ij}\Upsilon'\left(\frac{q_{ij}}{Q}\right)}{\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz} \tag{B.10}$$

We can now calculate how $Q$ changes when we consume $\bar{q}$ of $q_{ij}$ instead of $\underline{q}$. That is, the gain from consuming $q_{ij}$. This gain is given by

$$\int_{\underline{q}}^{\bar{q}}\frac{\partial Q}{\partial q_{ij}}dq_{ij}. \tag{B.11}$$

We can use the expression for $\frac{\partial Q}{\partial q_{ij}}$ to obtain:

$$\begin{aligned}
\int_{\underline{q}}^{\bar{q}}\frac{\partial Q}{\partial q_{ij}}dq_{ij} &= \int_{\underline{q}}^{\bar{q}}Q\frac{\tau_{ij}\Upsilon'\left(\frac{q_{ij}}{Q}\right)}{\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz}dq_{ij}\\
&= Q\frac{1}{\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz}\int_{\underline{q}}^{\bar{q}}\tau_{ij}\Upsilon'\left(\frac{q_{ij}}{Q}\right)dq_{ij}\\
&= Q\frac{1}{\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz}\tau_{ij}\;\Upsilon\left(\frac{q_{ij}}{Q}\right)\Bigg|_{\underline{q}}^{\bar{q}}\\
&= Q^2\frac{1}{\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz}\tau_{ij}\left[\Upsilon\left(\frac{\bar{q}}{Q}\right)-\Upsilon\left(\frac{\underline{q}}{Q}\right)\right]\\
&= \tau_{ij}QD\left[\Upsilon\left(\frac{\bar{q}}{Q}\right)-\Upsilon\left(\frac{\underline{q}}{Q}\right)\right],
\end{aligned}$$

where

$$D=\frac{Q}{\int\tau_{iz}\Upsilon'\left(\frac{q_{iz}}{Q}\right)q_{iz}dz}. \tag{B.12}$$

And for $\underline{q} = 0$, we obtain the desired expression. ■

We can then set $U(\tau)$ as follows,

$$U(\tau) = \tau QDP \left[ \Upsilon\left(\frac{q}{Q}\right) - \Upsilon(0) \right] - T,$$

so that

$$b(q, \tau) = \tau QDP \left[ \Upsilon\left(\frac{q}{Q}\right) - \Upsilon(0) \right].$$

This implies that

$$\begin{aligned}
\frac{\partial b(q,\tau)}{\partial q} &= \tau DP\Upsilon'\left(\frac{q}{Q}\right), \\
\frac{\partial^2 b(q,\tau)}{\partial \tau \partial q} &= DP\Upsilon'\left(\frac{q}{Q}\right)
\end{aligned}$$

The optimality condition (B.1) becomes

$$\tau D\Upsilon'\left(\frac{q}{Q}\right) = \frac{c}{P} + h(\tau)^{-1} D\Upsilon'\left(\frac{q}{Q}\right). \tag{B.13}$$

so that

$$\tau D\Upsilon'\left(\frac{q}{Q}\right) = \frac{c}{P}\frac{\tau}{\tau - h(\tau)^{-1}} \tag{B.14}$$

and

$$\frac{T(\tau)}{P} = \tau QD \left[ \Upsilon\left(\frac{q(\tau)}{Q}\right) - \Upsilon(0) \right] - QD \int_1^\tau \left[ \Upsilon\left(\frac{q(t)}{Q}\right) - \Upsilon(0) \right] dt \tag{B.15}$$

## B.3 Klenow-Willis specification

We use the Klenow-Willis formulation. That is,

$$\begin{aligned}
\Upsilon(q) &= 1 + (\sigma - 1)\exp\left(\frac{1}{\varepsilon}\right)\varepsilon^{\sigma/\varepsilon - 1}\left(\Gamma\left(\frac{\sigma}{\varepsilon}, \frac{1}{\varepsilon}\right) - \Gamma\left(\frac{\sigma}{\varepsilon}, \frac{q^{\varepsilon/\sigma}}{\varepsilon}\right)\right), && \text{(B.16)} \\
\Upsilon'(q) &= \frac{\sigma - 1}{\sigma} e^{\frac{1 - q^{\epsilon/\sigma}}{\epsilon}}, && \text{(B.17)} \\
\Upsilon''(q) &= -\frac{\sigma - 1}{\sigma} e^{\frac{1 - q^{\epsilon/\sigma}}{\epsilon}} \left(\frac{1}{\sigma} q^{\epsilon/\sigma - 1}\right), && \text{(B.18)}
\end{aligned}$$

We can invert $\Upsilon'(q)$ to

$$\left(\Upsilon'\right)^{-1}(x) = \left(1 - \epsilon \ln\left(\frac{\sigma}{\sigma - 1}x\right)\right)^{\frac{\sigma}{\epsilon}} \tag{B.19}$$

so that

$$\left(\frac{q}{Q}\right)^{\frac{\epsilon}{\sigma}} = 1 - \epsilon \ln\left(\frac{\sigma}{\sigma - 1}\right) - \epsilon \ln c + \epsilon \ln\left(\tau - h(\tau)^{-1}\right) + \epsilon \ln(PD) \tag{B.20}$$

Plugging in the Klenow-Willis formulation to the definition of $Q$ equation, we get

$$\int\int \tau\Gamma\left(\frac{\sigma}{\epsilon}, \frac{\left(\frac{q(c,\tau)}{Q}\right)^{\epsilon/\sigma}}{\epsilon}\right) dG(\tau)dF(c) = \int\int \tau\Gamma\left(\frac{\sigma}{\epsilon}, \frac{1}{\epsilon}\right) dG(\tau)dF(c) \tag{B.21}$$

### B.3.1 Second-best allocation

Consider a social planner that can set firm-level taxes and subsidies. Let's write the planner's problem of minimizing labor subject to providing a unit of aggregate consumption.

$$\min_{\{q(c,\tau),s(c)\}} \int\int cq(c,\tau)dG(\tau)dF(c) \tag{B.22}$$

$$\text{s.t.} \qquad \tau\Upsilon'\left(q(c,\tau)\right) = \frac{\tau}{\tau - h(\tau)^{-1}}c(1-s(c)), \qquad [\gamma(c,\tau)] \tag{B.23}$$

$$\int\int \tau\Upsilon\left(q(c,\tau)\right) dG(\tau)dF(c) = 1, \qquad [\mu] \tag{B.24}$$

Take FOC:

$$[s(c)]: \qquad \int \frac{\tau}{\tau - h(\tau)^{-1}}\gamma(c,\tau)dG(\tau) = 0, \tag{B.25}$$

$$[q(c,\tau)]: \qquad c = \tau\Upsilon''(q(c,\tau))\gamma(c,\tau) + \tau\Upsilon'(q(c,\tau))\mu \tag{B.26}$$

**Numerical algorithm to compute the second-best allocation.** Below we outline a sketch of the algorithm we use to compute the second-best allocation:

1. Guess Lagrange multiplier $\mu$.
2. For each cost level $c$:
   (a) Guess the subsidy of firms with such cost level, $s(c)$
   (b) Use (B.23) to to obtain $q(c,\tau)$
   (c) Use (B.26) to obtain $\gamma(c,\tau)$
   (d) Check if (B.25) holds
   (e) Iterate until finding $s(c)$
3. Use the constraint (B.24) as a residual equation.
4. Iterate until correct $\mu$ is found and residual equation is approximately zero.

## B.4 Quantitative analysis

The model moments and calibrated parameters are displayed in Table B.1.

The misallocation costs are summarized in Table B.2. The welfare costs from misallocation with Kimball preferences are similar to our benchmark results: 0.79% relative to 0.68%.

Table B.1: Calibrated Moments and Parameters under Kimball

| Parameter | | Model | | Moment | | Model | |
|---|---|---|---|---|---|---|---|
| | | Benchm. | Kimball | | Data | Benchm. | Kimball |
| $\bar{\tau}$ | Highest taste | 1.19 | 1.21 | Package size dispersion | 0.44 | 0.44 | 0.44 |
| $\theta$ | Pareto shape | 3.14 | 3.18 | Sales share top 5% | 0.73 | 0.73 | 0.63 |
| $\eta$ | Elasticity of differences | 4.15 | – | Aggregate markup | 1.3 | 1.3 | 1.32 |
| $\beta_0$ | Departure from CES | 0.02 | – | Markup elast. w.r.t. size | 0.03 | 0.03 | 0.03 |
| $\sigma$ | Degree of demand elasticity | – | 5.38 | | | | |
| $\epsilon$ | Degree of superelasticity | – | 0.87 | | | | |

Table B.2: Welfare Costs of Misallocation

| Baseline Model | Kimball Specification |
|---|---|
| 0.68% | 0.79% |

**Notes:** This table reports the welfare costs in the decentralized equilibrium relative to the efficient allocation for our benchmark model (column 1), and the model with Kimball preferences (column 2). All welfare costs are measured in consumption equivalent terms—that is, the uniform decline in consumption that would make households indifferent between the two equilibria.

Proposition 7 shows that under CED, the planner has no incentive to use firm-level taxes and subsidies. Since Kimball preferences do not satisfy CED, the planner can potentially reduce misallocation by imposing firm-level taxes and subsidies. We find that quantitatively, the planner can improve welfare by a very modest amount. The optimal allocation with firm-level taxes and subsidies only improves the decentralized equilibrium by 0.02% in consumption equivalent units. That is, it eliminates a very small portion of the overall welfare costs due to misallocation (0.79%).

# C Imperfect substitution within firms

In the benchmark model, firms offer a quantity to be sold to low-taste consumers and a quantity to high-taste consumers. Recall that the utility of household $i$ from purchasing quantity $q_{ij}$ is given by

$$U_{ij} = \tau_{ij} u(q_{ij}),$$

where $\tau_{ij}$ is the taste of household $i$ for the product of firm $j$. Because $q_{ij}$ specifies the *total* amount consumed by household $i$, purchasing multiple small-sized packages is perfectly substitutable with consuming a single large-sized package.

As we explain in detail below, the *perfect substitutes assumption* does not drive our two main results: (i) nonlinear pricing can lead to misallocation across consumers within a firm, (ii) firm-level markups may not be informative of the degree of misallocation across firms. The key assumption in our setup that underlies the difference relative to standard models of linear pricing is that goods are *indivisible*, that is, firms can mandate a minimum quantity sold (package size) together with price. Consumers cannot purchase fractions of large units.

Expanding our environment to incorporate imperfectly substitutable goods is not straightforward. Typical definitions of imperfect substitutability apply to pre-specified goods (e.g., apples vs. oranges or Coca-Cola vs. Pepsi), rather than different quantities of the same good, where the quantities are chosen by firms.

In this section, we consider three alternative ways of incorporating imperfect substitution across the same product sold in a variety of sizes. The first preserves the choice of firms and the social planner of what a "large" and "small" package is, but allows consumers to purchase multiple bundles of the same firm. In this specification, we assume consumers' utility is not only a function of the total quantity consumed, but also depends on how the quantities are purchased, i.e., in which package sizes. The second and third models consider environments where units are chosen by nature, allowing us to connect to more standard models of imperfect substitutability. In the second model, we maintain the discrete choice nature of the problem, but add an idiosyncratic taste towards every specific sized bundle. In the third specification, we allow consumers to purchase multiple bundles and assume there is a constant elasticity of substitution across the different sized bundles. In all three, we show that our two main results continue to hold.

## C.1 Disutility of package size

Suppose that consumer utility not only depends on the total quantity of the good purchased, but also on *how* it is purchased, for example, the way in which goods are packaged. Let $q_s$, $s \in \{1, ..., S\}$ be the physical quantity of the good contained in each available size $s$, and let $n_s$ be the number of packages of each size purchased.

Utility now has three components.

$$\tau u \left( \sum_s n_s \times q_s \right) - \sum_s n_s \times v(q_s) - f \left( \sum_s n_s \right) \tag{C.1}$$

The first component is the same as in the main text and represents the utility flow from physical quantity; say, fluid ounces of Coca Cola. The second component, $v(q_s)$, is an increasing function of package size that captures any disutility from the *way in which* the fluid ounces are purchased, e.g., disutility of carrying a large container, or more rapid depreciation once the package has been opened.[58] The third component captures a fixed cost of purchasing or opening a package.[59]

Taking into account the disutility associated with buying large vs small units, the two are no longer perfect substitutes, as long as $2v(q_l) - f(2) \neq v(2q_l) - f(1)$.

**Social planner's problem.** Given an aggregate price index $P^{\text{FB}}$, the planner chooses a set of sizes for each consumer type $\tau$ to maximize (C.1). For simplicity, normalize $f(1) = 0$ and assume that $f(2)$ is large enough such that the planner chooses to deliver the first-best quantities to each type in a single package.

The first-best allocations then solve

$$\tau_{ij} u'(q_{ij}^{FB}) - v'(q_{ij}^{FB}) = \frac{c_j}{P^{FB}} \tag{C.2}$$

**Market allocation.** The firm's problem has the same objective as in the main text, but the IR and the IC constraints are now different:

$$\begin{aligned}
\max_{\{q_{1j}, q_{\tau j}, p_{1j}, p_{\tau j}\}} \quad & \pi q_{\tau j}\left(p_{\tau j} - c_j\right) + (1-\pi) q_{1j}\left(p_{1j} - c_j\right) && \text{(C.3)} \\
\text{s.t} \quad & u(q_{1j}) - v(q_{1j}) - \frac{p_{1j} q_{1j}}{P} = 0 && [IR_1] \\
& \tau u(q_{\tau j}) - v(q_{\tau j}) - \frac{p_{\tau j} q_{\tau j}}{P} = \tau u(q_{1j}) - v(q_{1j}) - \frac{p_{1j} q_{1j}}{P} && [IC_\tau]
\end{aligned}$$

Note that we omitted the possibility that the consumer may want to purchase two small bundles; i.e. $f(2)$ is large enough.[60]

The quantities sold the high and low type respectively solve

$$\tau u'(q_{\tau j}) - v'(q_{\tau j}) = \frac{c_j}{P} \tag{C.4}$$

$$\frac{1-\tau\pi}{1-\pi} u'(q_{1j}) - v'(q_{1j}) = \frac{c_j}{P} \tag{C.5}$$

Similarly to the baseline model, any distortion at the top only comes from the difference in price indices between the planner and the market. The allocation to the low type features the same wedge as before, but the wedge only affect the first part of utility, since we assumes that the preference shifter does not affect utility of package size.[61].

[58] Here, the taste shifter only applies to the first component of utility. If we assumed that it also applies to $v(.)$, the problem becomes identical to the one we studied in the main text, with a utility function given by $u(q) - v(q)$.

[59] This is necessary to ensure that the planning problem has an interior solution for the set of $q_s$. Otherwise, it may be optimal to sell infinitely many packages of infinitesimal size.

[60] Allowing for a lower value of $f(2)$ such that consumers may want to purchase two small bundles would not change the tow main propositions below, but would considerably complicate notation.

[61] If the taste shifter also applied to $v(.)$, the problem would fully collapse to the original formulation

**Misallocation.** The following propositions show that the two main results of our baseline model with perfectly substitutable sizes continue to hold. The first mirrors Proposition 3. In general equilibrium, the allocation sold to both the high taste and the low taste consumer are distorted. The second mirrors Propositions 7 and 9: firms may charge size-dependent markups despite the fact that the allocation of production across firms is efficient.

PROPOSITION **19.** *Given a level of labor $l_j$, households consume too much of the goods for which they have a high taste and too little of the goods for which they have a low taste.*

PROOF OF PROPOSITION **19.** Given that $l_j$ is assumed to be constant, there are three possibilities: (1) $q_{1j} = q_{1j}^{FB}$ and $q_{\tau j} = q_{\tau j}^{FB}$, (2) $q_{1j} > q_{1j}^{FB}$ and $q_{\tau j} < q_{\tau j}^{FB}$, or (3) $q_{1j} < q_{1j}^{FB}$ and $q_{\tau j} > q_{\tau j}^{FB}$.

Suppose $q_{1j} = q_{1j}^{FB}$ and $q_{\tau j} = q_{\tau j}^{FB}$. From $q_{\tau j} = q_{\tau j}^{FB}$, it then follows that $P = P^{FB}$. Using (C.2) and (C.5), $q_{1j} = q_{1j}^{FB}$ would imply that $\frac{1-\tau\pi}{1-\pi} = 1$.

Suppose that $q_{1j} > q_{1j}^{FB}$ and $q_{\tau j} < q_{\tau j}^{FB}$. From $q_{\tau j} = q_{\tau j}^{FB}$, it then follows that $P < P^{FB}$. Using (C.2) and (C.5), we would have $q_{1j} < q_{1j}^{FB}$, contradicting this case as well. ■

The intuition behind Proposition 19 is similar to the baseline model. The additional utility cost of purchasing in larger packages applies to both types of consumers, in the planner's problem as well as in the market allocation. This cost in and of itself does not create any additional distortion. As in the baseline model, firms distort the quantity sold to low-taste consumers downward in order to be able to extract more rents from the top. For the labor market to clear, the aggregate price index must therefore increase, introducing a distortion also at the top.

In general, markup variation across firms as well as any patterns of misallocation of production depend on the shapes of $u'()$ and $v'()$. However, we can still show that our main result of no misallocation with CED preferences holds; as long as we extend the definition of CED to also include the disutility of package size.

PROPOSITION **20.** *Markup variation across firms is not necessarily a sign of misallocation across firms. That is, there exist utility functions $u(.)$ and disutility functions $v(.)$ such that there is markup variation across firms, but no misallocation.*

PROOF OF PROPOSITION **20.** Suppose that the disutility of package size $v$ is proportional to $u$, that is, $v(q_{ij}) = \nu u(q_{ij})$ for some $\nu < 1$. Intuitively, this would mean that one looses $\nu$% of utility from, say, Coke getting flat after opening a large bottle. Then, if $\tau u(q)$ features CED, so does $\tau u(q) - v(q) = (\tau - \nu)v(q)$ and there is no misallocation of labor across firms in equilibrium. ■

Intuitively, when the entire utility of purchasing and consuming a certain package size features CED, the disutility of buying a big bottle of Coca Cola acts as a proportional downward shift that affects utility of all consumers proportionally. The model with imperfect substitutes then is analogous to a model in which different sizes are perfect substitutes, but there is *less* taste dispersion across households: low-taste households have effective taste shifters of $(1-\nu)$, while the taste shifter is $\tau - \nu$ at the top.

## C.2 Continuum of tastes and fixed sizes

Now consider an environment where Households consume a variety of product types, $j \in (0,1)$, but where each product type is offered in two *exogenous* sizes, $q_{jl}$ and $q_{jh}$. The preferences of household $i$ are given by

$$U_i = \int_0^1 \left[\mathbb{I}(s_{ij} = l)\left(\tau_{ij} u(q_{jl})\right) + \mathbb{I}(s_{ij} = h)\left(\tau_{ij} u(q_{jh})\right)\right] dj, \tag{C.6}$$

where $s_j \in \{n, l, h\}$ indicates whether the consumer chooses to purchase no good ($n$), the low-quantity bundle ($l$) or the high-quantity bundle ($h$). $\tau_{ij}$ is the idiosyncratic taste of household $i$ for good $j$ and is uniformly distributed between $[1, \tau]$ with $\tau < 2$.

**Firm's problem.** Let $\underline{\tau}$ be the threshold levels below which consumers do not purchase either bundle and $\bar{\tau}$ be the threshold above which consumers buy the high bundle. The firm's problem (multiplied by $(\tau - 1)$) is given by

$$\begin{aligned}
\max_{\{p_l, p_h, \underline{\tau}, \bar{\tau}\}} \quad & (p_l - c_j) q_l (\bar{\tau} - \underline{\tau}) + (p_h - c_j) q_h (\tau - \bar{\tau}) \\
\text{s.t.} \quad & \underline{\tau} u(q_l) = \frac{p_l q_l}{P}, \\
& \bar{\tau} u(q_h) - \frac{p_h q_h}{P} = \bar{\tau} u(q_l) - \frac{p_l q_l}{P}.
\end{aligned}$$

Taking FOCs,

$$[\underline{\tau}]: \quad (p_l - c_j) q_l = \lambda_1 u(q_l) \tag{C.7}$$

$$[\bar{\tau}]: \quad (p_h - c_j) q_h - (p_l - c_j) q_l = \lambda_2 \left(u(q_h) - u(q_l)\right) \tag{C.8}$$

$$[p_l]: \quad q_l (\bar{\tau} - \underline{\tau}) = \lambda_1 \frac{q_l}{P} - \lambda_2 \frac{q_l}{P} \tag{C.9}$$

$$[p_h]: \quad q_h (\tau - \bar{\tau}) = \lambda_2 \frac{q_h}{P} \tag{C.10}$$

From the FOC wrt $p_h$ and $p_l$ we get

$$\lambda_2 = P(\tau - \bar{\tau}) \tag{C.11}$$

$$\lambda_1 = P(\tau - \underline{\tau}) \tag{C.12}$$

Combining the FOC w.r.t. $\underline{\tau}$ with the expression for $\lambda_1$, we get

$$\frac{(p_l - c_j) q_l}{P} = (\tau - \underline{\tau})\, u(q_l).$$

Using the IR constraint,

$$\underline{\tau} u(q_l) - \frac{c_j q_l}{P} = (\tau - \underline{\tau})\, u(q_l)$$

so that

$$\underline{\tau} = \frac{\tau}{2} + \frac{1}{2} \frac{c_j}{P} \frac{q_l}{u(q_l)}. \tag{C.13}$$

Now using the FOC w.r.t. $\bar{\tau}$, combining with the expression for $\lambda_2$, we get

$$\frac{p_h q_h - p_l q_l}{P} - (q_h - q_l)\frac{c_j}{P} = (\tau - \bar{\tau})\left(u(q_h) - u(q_l)\right),$$

Using the IC constraint, we obtain

$$\bar{\tau}(u(q_h) - u(q_l)) - (q_h - q_l)\frac{c_j}{P} = (\tau - \bar{\tau})\left(u(q_h) - u(q_l)\right),$$

So that

$$\bar{\tau} = \frac{\tau}{2} + \frac{1}{2}\frac{q_h - q_l}{u(q_h) - u(q_l)}\frac{c_j}{P} \tag{C.14}$$

We assume that $q_h, q_l, u(q_h), u(q_l)$ are such that $1 < \underline{\tau}$ and $\bar{\tau} < \tau$ for all firms. We can then obtain the prices from the IR and IC constraints:

$$p_l = \underline{\tau}\frac{u(q_l)}{q_l}P,$$
$$p_h = \bar{\tau}\frac{u(q_h) - u(q_l)}{q_h}P + p_l\frac{q_l}{q_h}.$$

**Misallocation.** In this setup, the source of misallocation across consumers is governed by the two thresholds $\underline{\tau}$ and $\bar{\tau}$. Holding constant the overall production of a firm, a social planner may choose to give low taste consumers who fall below $\underline{\tau}$ the small packaged product at the expense of higher taste consumers who will obtain the small instead of the large packaged good. The next proposition formalizes this form of misallocation.

PROPOSITION **21.** *If a social planner chooses how to allocate the production of a firm to different consumers, they would choose a lower $\underline{\tau}$ and a higher $\bar{\tau}$. That is, they would serve more low-taste customers at the expense of downsizing the bundle of some high-taste consumers.*

Proof. The planner's problem is

$$\max_{\{\underline{\tau},\bar{\tau}\}} \quad \frac{1}{2}(\bar{\tau}^2 - \underline{\tau}^2)u(q_l) + \frac{1}{2}(\tau^2 - \bar{\tau}^2)u(q_h)$$
$$\text{s.t.} \qquad (\bar{\tau} - \underline{\tau})q_l + (\tau - \bar{\tau})q_h = \frac{L_j}{c_j}.$$

Taking FOC:

$$[\underline{\tau}]: \qquad \underline{\tau}u(q_l) = q_l\lambda,$$
$$[\bar{\tau}]: \qquad \bar{\tau}\left(u(q_h) - u(q_l)\right) = (q_h - q_l)\lambda,$$

So that

$$\frac{\bar{\tau}^*}{\underline{\tau}^*} = \frac{q_h - q_l}{q_l}\frac{u(q_l)}{u(q_h) - u(q_l)} \tag{C.15}$$

While the decentralized one has

$$\frac{\bar{\tau} - \frac{\tau}{2}}{\underline{\tau} - \frac{\tau}{2}} = \frac{q_h - q_l}{q_l} \frac{u(q_l)}{u(q_h) - u(q_l)} \tag{C.16}$$

That is, for the same level of $\underline{\tau}$, the planner would choose $\bar{\tau}^* > \bar{\tau}$. Since that wouldn't exhaust the resources, it must be that $\underline{\tau}^* < \underline{\tau}$. Similarly, for the same level of $\bar{\tau}$, the planner would choose a $\underline{\tau}^* < \underline{\tau}$. Since that allocation is not feasible with the firm's employment, it must be that $\bar{\tau}^* > \bar{\tau}$. ∎

Now suppose the economy is populated with two types of firms—a low productivity ($c_1$) and a high productivity one ($c_2 < c_1$). The small and large packages sizes are also firm specific. Consider a planner that can impose taxes and subsidies at the firm level. The planner's problem is

$$\begin{aligned}
\max_{\{\underline{\tau}_1,\bar{\tau}_1,\underline{\tau}_2,\bar{\tau}_2\}} \quad & \frac{1}{2}(\bar{\tau}_1^2 - \underline{\tau}_1^2)u(q_{l1}) + \frac{1}{2}(\tau^2 - \bar{\tau}_1^2)u(q_{h1}) + \frac{1}{2}(\bar{\tau}_2^2 - \underline{\tau}_2^2)u(q_{l2}) + \frac{1}{2}(\tau^2 - \bar{\tau}_2^2)u(q_{h2}) \\
\text{s.t.} \quad & \left[(\bar{\tau}_1 - \underline{\tau}_1)q_{l1} + (\tau - \bar{\tau}_1)q_{h1}\right] c_1 + \left[(\bar{\tau}_2 - \underline{\tau}_2)q_{l2} + (\tau - \bar{\tau}_2)q_{h2}\right] c_2 = L, \\
& \bar{\tau}_1 = -\left[\frac{q_{h1}u(q_{l1}) - q_{l1}u(q_{h1})}{q_{l1}\left(u(q_{h1}) - u(q_{l1})\right)}\right]\frac{\tau}{2} + \frac{u(q_{l1})}{q_{l1}}\frac{q_{h1} - q_{l1}}{u(q_{h1}) - u(q_{l1})}\underline{\tau}_1 \\
& \bar{\tau}_2 = -\left[\frac{q_{h2}u(q_{l2}) - q_{l2}u(q_{h2})}{q_{l2}\left(u(q_{h2}) - u(q_{l2})\right)}\right]\frac{\tau}{2} + \frac{u(q_{l2})}{q_{l2}}\frac{q_{h2} - q_{l2}}{u(q_{h2}) - u(q_{l2})}\underline{\tau}_2
\end{aligned}$$

We now show that the firm-level markup is not indicative of whether such firm is too small or too big. That is, the relative markup of a firm cannot be used to determined if such firm should be subsidized or taxed.

Proposition **22.** *Markup variation across firms is not a sufficient statistic for misallocation across firms. That is, there exist parameters such that the planner would optimally subsidize some firms at the expense of others, even though such firms charge a lower markup. And, similarly, there exist parameters such that the planner would optimally subsidize some firms at the expense of others, when such firms charge a higher markup.*

Proof.

Taking first order conditions

$$\begin{aligned}
[\underline{\tau}_1]: \quad & \underline{\tau}_1 u(q_{l1}) = q_{l1}c_1\lambda + \frac{u(q_{l1})}{q_{l1}}\frac{q_{h1} - q_{l1}}{u(q_{h1}) - u(q_{l1})}\nu_1, \\
[\underline{\tau}_2]: \quad & \underline{\tau}_2 u(q_{l2}) = q_{l2}c_2\lambda + \frac{u(q_{l2})}{q_{l2}}\frac{q_{h2} - q_{l2}}{u(q_{h2}) - u(q_{l2})}\nu_2, \\
[\bar{\tau}_1]: \quad & \bar{\tau}_1\left(u(q_{h1}) - u(q_{l1})\right) = (q_{h1} - q_{l1})c_1\lambda - \nu_1, \\
[\bar{\tau}_2]: \quad & \bar{\tau}_2\left(u(q_{h2}) - u(q_{l2})\right) = (q_{h2} - q_{l2})c_2\lambda - \nu_2,
\end{aligned}$$

Combining the first order conditions we obtain

$$\underline{\tau}_i u(q_{li}) = q_{li}c_i\lambda + \frac{u(q_{li})}{q_{li}}\frac{q_{hi} - q_{li}}{u(q_{hi}) - u(q_{li})}\left[(q_{hi} - q_{li})c_i\lambda - \bar{\tau}_i\left(u(q_{hi}) - u(q_{li})\right)\right] \tag{C.17}$$

for $i \in \{1,2\}$. Rearranging and plugging the value for $\bar{\tau}_i$:

$$\underline{\tau}_i u(q_{li}) = \left[ q_{li} + \frac{u(q_{li})}{q_{li}} \frac{(q_{hi}-q_{li})^2}{u(q_{hi})-u(q_{li})} \right] c_i \lambda$$
$$- \frac{u(q_{li})}{q_{li}}(q_{hi}-q_{li}) \left[ - \left[ \frac{q_{hi}u(q_{li}) - q_{li}u(q_{hi})}{q_{li}\left(u(q_{hi})-u(q_{li})\right)} \right] \frac{\tau}{2} + \frac{u(q_{li})}{q_{li}} \frac{q_{hi}-q_{li}}{u(q_{hi})-u(q_{li})} \underline{\tau}_i \right]. \tag{C.18}$$

We can rearrange this equation to be

$$D_i \underline{\tau}_i = A_i + B_i c_i, \tag{C.19}$$

with

$$A_i = \frac{u(q_{li})}{q_{li}} \frac{q_{hi}-q_{li}}{u(q_{hi})-u(q_{li})} \frac{q_{hi}u(q_{li}) - q_{li}u(q_{hi})}{q_{li}} \frac{\tau}{2} \tag{C.20}$$
$$D_i = u(q_{li}) \left[ 1 + \frac{u(q_{li})}{q_{li}} \frac{q_{hi}-q_{li}}{u(q_{hi})-u(q_{li})} \frac{q_{hi}-q_{li}}{q_{li}} \right] \tag{C.21}$$

where $B_i$'s formula is irrelevant for the rest of the proof so we omit it. We have that

$$\frac{\underline{\tau}_2 - A_2/D_2}{\underline{\tau}_1 - A_1/D_1} = \frac{c_2}{c_1} \tag{C.22}$$

Recall that in the market equilibrium we have

$$\frac{\underline{\tau}_2 - \frac{\tau}{2}}{\underline{\tau}_1 - \frac{\tau}{2}} = \frac{c_2}{c_1}. \tag{C.23}$$

We confirm numerically that different sets of $\{q_{l1}, q_{l2}, q_{h1}, q_{h2}, \tau, c_1, c_2, u(\cdot)\}$ deliver different implications for the direction of subsiddies relative to markups.

■

## C.3 Nested CES preferences

The last model we consider is one with nested CES preferences as in Christian Broda and David E. Weinstein (2010). Let $\sigma$ denote the elasticity of substitution between different package sizes sold by the same firm, while $\gamma$ is the elasticity of substitution across firms.

As in the example above, nature chooses two package sizes: $q_h$ and $q_l$, and $n_h$ and $n_l$ denote the number of packages consumed. There is a unit continuum of identical firms in the economy that produce with a linear technology and unit cost normalized to one.

The utility of purchasing $n_h$ units of the large package $q_h$ and $n_l$ units of the small package $q_l$ from firm $j$ is given by

$$U_i = \int_0^1 \tau_{ij} \left( \left( (n_{h,ij} q_h)^{\frac{\sigma-1}{\sigma}} + (n_{l,ij} q_l)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\gamma-1}{\gamma}} dj \tag{C.24}$$

For simplicity, suppose that there are two types, $\tau_{ij} = 1$ and $\tau_{ij} = \tau > 1$ and there are equal shares of both types in the population.

**Social planner allocation.** Suppose labor is scarce so that the planner can produce one unit of $q_h$ and one unit of $q_l$ per firm. Suppose further that the optimal allocation features $\{(1,0),(0,1)\}$, meaning that consumers receive 1 unit of $q_l$ for all the goods for which they have a low taste and 1 unit of $q_h$ for all the goods for which they have a low taste.[62]

**Market allocation.** Each firm chooses a unit price $p_h$ and $p_l$ as well as an allocation—units sold to high and low type—to maximize profits. We focus on symmetric equilibria. That is, each firm uses the same amount of labor and hence decides between the same allocations as the planner: $\{(1,0),(0,1)\}$ or $\{(0,0),(1,1)\}$. We omit the $j$ subscript from now on.

If the firm chooses $\{(1,0),(0,1)\}$, it solves the following problem[63]

$$\max_{\{p_l,p_h\}} \quad q_h\left(p_h - c\right) + q_l\left(p_l - c\right) \tag{C.25}$$

$$\text{s.t} \quad q_l^{\frac{\gamma-1}{\gamma}} - \frac{p_l q_l}{P} = 0, \qquad [IR_1]$$

$$\tau q_h^{\frac{\gamma-1}{\gamma}} - \frac{p_h q_h}{P} = \tau q_l^{\frac{\gamma-1}{\gamma}} - \frac{p_l q_l}{P}. \qquad [IC_\tau]$$

If the firm chooses $\{(0,0),(1,1)\}$, it excludes the low taste consumer and charges the high taste consumer a transfer T in exchange for the full bundle.

$$\max_{\{T\}} \quad T - c(q_h + q_l) \tag{C.26}$$

$$\text{s.t.} \quad \tau\left(\left(q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\gamma-1}{\gamma}} \le \frac{T}{P}.$$

The profits of the two options are given by

$$\Pi(\{(1,0),(0,1)\}) = P\left[\tau q_h^{\frac{\gamma-1}{\gamma}} + (2-\tau)\, q_l^{\frac{\gamma-1}{\gamma}}\right], \tag{C.27}$$

$$\Pi(\{(0,0),(1,1)\}) = P\tau\left[\left(q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right]^{\frac{\gamma-1}{\gamma}}. \tag{C.28}$$

When choosing to exclude the low type, the firm can extract the full consumer surplus from the

[62] It is easy to show that the opposite allocation – $q_h$ to low type, $q_l$ to high type – is strictly worse and that giving both $q_h$ and $q_l$ to the high type is always worse than giving that same allocation to the high type. As long as $\tau q_h^{\frac{\gamma-1}{\gamma}} + q_l^{\frac{\gamma-1}{\gamma}} \ge \tau\left(\left(q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\gamma-1}{\gamma}}$, it is optimal for the planner to choose $\{(1,0),(0,1)\}$ over $\{(0,0),(1,1)\}$.

[63] Profits are multiplied by 2 for ease of readability.

high type, but when serving both, the IC constraints ensures that the high type receives an information rent.

**Misallocation.** We start by showing that the market allocation may feature misallocation of consumption across consumers, where high-taste consumers consume too much of the good and low-taste consumers too little.

PROPOSITION **23.** *There exist parameters such that households consume too much of the goods for which they have a high taste and too little of the goods for which they have a low taste.*

PROOF OF PROPOSITION **23.** This case happens whenever the planner chooses to serve both types, but firms exclude low taste consumers and sell both packages to the high taste consumer in order to extract the full surplus.

We construct one such example below. Let $\varepsilon \geq 0$ be the difference between the social value of serving both types and the one of providing both $q_h$ and $q_l$ to the high type.

$$\varepsilon \equiv \tau q_h^{\frac{\gamma-1}{\gamma}} + q_l^{\frac{\gamma-1}{\gamma}} - \tau \left( \left( q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\gamma-1}{\gamma}} \tag{C.29}$$

The difference between firm profits from serving both types and only the high taste consumer can be written as

$$\frac{\Pi(\{(1,0),(0,1)\}) - \Pi(\{(0,0),(1,1)\})}{P} = \tau q_h^{\frac{\gamma-1}{\gamma}} + q_l^{\frac{\gamma-1}{\gamma}} - (\tau-1)\, q_l^{\frac{\gamma-1}{\gamma}} - \tau \left[ \left( q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \right]^{\frac{\gamma-1}{\gamma}} \tag{C.30}$$

$$= \varepsilon - (\tau-1)\, q_l^{\frac{\gamma-1}{\gamma}}, \tag{C.31}$$

which is negative for small enough $\varepsilon$. That, is, there are parameters such that the social planner chooses to allocate some of all goods to all people $\varepsilon \geq 0$, but the market allocates too much ($q_l$ in addition to $q_h$) to the high type and too little (nothing) to the low type. ■

Finally, we show that also in this model, the relative firm-level markup is not indicative of whether a social planner would prefer to subsidize or tax the firm.

PROPOSITION **24.** *Markup variation across firms is not a sufficient statistic for misallocation across firms. That is, there exist parameters such that the planner would optimally subsidize some firms at the expense of others, even though both firms charge the same markup.*

PROOF OF PROPOSITION **24.** In the symmetric equilibrium, all firms charge the same markup. It remains to be shown that the social planner may want to re-allocate labor across firms. Suppose the social planner moves $q_l$ workers from one firm to another. The firm that now only employs $q_h$ workers sells that package to high types, which results in a welfare loss of

$$\text{Welfare loss} = \tau\left(\left(q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\gamma-1}{\gamma}} - \tau q_h^{\frac{\gamma-1}{\gamma}} \tag{C.32}$$

The firm that now employs $q_h + 2q_l$ workers can either sell all three units to the high type and extract the full surplus, or sell the additional unit to the low type, extract the full surplus there and reduce the transfer charged to the high type. We verify numerically that there are parameter values such that the firm chooses to sell the extra unit to the low type, while preferring $\{(0,0)(1,1)\}$ to $\{(1,0)(0,1)\}$ as before.

The welfare gains from the firm employing an additional $q_l$ workers are given by

$$\text{Welfare gain} = q_l^{\frac{\gamma-1}{\gamma}} \tag{C.33}$$

Welfare gains exceeds losses whenever

$$q_l^{\frac{\gamma-1}{\gamma}} + \tau q_h^{\frac{\gamma-1}{\gamma}} \geq \tau\left(\left(q_h^{\frac{\sigma-1}{\sigma}} + q_l^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\gamma-1}{\gamma}}, \tag{C.34}$$

which is precisely the maintained assumption under which serving both types is socially preferable to allocation all the good to the high taste consumer.

■

# D Oligopolistic competition

In this section, we consider a modification to the baseline model whereby differences in market power of firms come not from consumer preferences, but from the market structure. In particular, we depart from monopolistic competition and set up an environment in which firms compete a la Atkeson and Burstein (2008).

**Setup.** There is a unit mass of product types $m \in (0, 1)$. In each type, there are two firms, 1 and 2, that have a linear production technology with unit costs $c_1 < c_2$ respectively. Consumers have nested CES preferences over firms and sectors: within product type, the goods produced by each firm have an elasticity of substitution $\epsilon$, while across sectors, good are more substitutable, with an elasticity $\sigma > \epsilon > 1$.

As in the baseline model, we allow for taste heterogeneity across consumers. Consumers have i.i.d. tastes $\tau_{im}$ towards each of the product types, where $\tau_{im}$ can take on two values, $\tau$ and 1. Consumers supply one unit of labor inelastically and own the firms.

Preferences of consumer $i$ are given by

$$U_i = \int_0^1 \tau_{im} q_{im}^{\frac{\sigma-1}{\sigma}} dm, \tag{D.1}$$

$$q_{im} = \left( q_{ijm}^{\frac{\epsilon-1}{\epsilon}} + q_{ikm}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{D.2}$$

**Firm's problem.** We will consider the problem of firm $j \in \{1, 2\}$ who competes with the other firm, $-j$. Similar to our benchmark environment, one can easily show that the IC constraint does not bind for the low-taste consumer. For now, let us assume that the IC constraint of the high-taste consumer is binding. In this case, the firm's problem is given by

$$\max_{\{p_{jl}, p_{jh}, q_{jl}, q_{jh}\}} \pi(p_{jh} - c_j) q_{jh} + (1 - \pi) (p_{jl} - c_j) q_{jl}, \tag{D.3}$$

$$\text{s.t.} \quad \left( q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} = (q_{-jl})^{\frac{\sigma-1}{\sigma}} + \frac{p_{jl} q_{jl}}{P}, \tag{D.4}$$

$$\tau \left( q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jh}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} = \tau \left( q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} + \frac{p_{jh} q_{jh} - p_{jl} q_{jl}}{P}. \tag{D.5}$$

Taking first order conditions:

$$
\begin{aligned}
[p_{jh}]: \qquad & \pi = \frac{1}{P}\lambda_{jh}, \\
[p_{jl}]: \qquad & (1-\pi) = \frac{1}{P}\left(\lambda_{jl} - \lambda_{jh}\right), \\
[q_{jh}]: \qquad & \pi(p_{jh} - c_{jh}) = \left[-\frac{\sigma-1}{\sigma}\tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jh}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jh}^{-\frac{1}{\epsilon}} + \frac{p_{jh}}{P}\right]\lambda_{jh}, \\
[q_{jl}]: \qquad & (1-\pi)(p_{jl} - c_{jl}) = \left[-\frac{\sigma-1}{\sigma}\left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}} + \frac{p_{jl}}{P}\right]\lambda_{jl} \\
& \qquad - \left[-\frac{\sigma-1}{\sigma}\tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}} + \frac{p_{jl}}{P}\right]\lambda_{jh},
\end{aligned}
$$

From the first two equations, we obtain:

$$
\begin{aligned}
\lambda_{jh} &= \pi P, \\
\lambda_{jl} &= P.
\end{aligned}
$$

The high quantity optimality condition then becomes

$$
\frac{c_{jh}}{P} = \frac{\sigma-1}{\sigma}\tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jh}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jh}^{-\frac{1}{\epsilon}}. \tag{D.6}
$$

The low quantity optimality condition can be rearranged to

$$
(1-\pi)\frac{c_{jl}}{P} = \frac{\sigma-1}{\sigma}\left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}} - \pi\frac{\sigma-1}{\sigma}\tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}}
$$

or

$$
\frac{\sigma-1}{\sigma}\left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}} = \frac{1-\pi}{1-\pi\tau\psi_j}\frac{c_j}{P} \tag{D.7}
$$

where $\psi_j \in (0,1)$ is given by

$$
\psi_j = \left(\frac{q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}}{q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1}. \tag{D.8}
$$

Recall that we have assumed that the IC constraint rather than the IR constraint is binding. Note that when both IR constraints are binding, the allocations and prices are given by

$$\frac{c_j}{P} = \frac{\sigma-1}{\sigma}\tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jh}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jh}^{-\frac{1}{\epsilon}}, \tag{D.9}$$

$$\frac{c_j}{P} = \frac{\sigma-1}{\sigma}\tau\left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}}, \tag{D.10}$$

$$\frac{p_{jh}q_{jh}}{P} = \tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jh}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} - \tau q_{-jh}^{\frac{\sigma-1}{\sigma}}, \tag{D.11}$$

$$\frac{p_{jl}q_{jl}}{P} = \left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} - q_{-jl}^{\frac{\sigma-1}{\sigma}}, \tag{D.12}$$

To check whether the IR allocations satisfies the IC constraint, we need to check the following condition:

$$\tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jh}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} \geq \tau\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} + \frac{p_{jh}q_{jh} - p_{jl}q_{jl}}{P}$$

Rearranging with the IR pricing:

$$\tau\left[q_{-jh}^{\frac{\sigma-1}{\sigma}} - \left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}}\right] + \left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} - q_{-jl}^{\frac{\sigma-1}{\sigma}} \geq 0 \tag{D.13}$$

If this condition holds, the equilibrium allocation is given by the IR allocation. Otherwise, it is given by the IC allocation (D.7).

Lemma **4.** *If the in the market equilibrium the incentive compatibility constraint is binding, it must be that $\psi_j > \frac{1}{\tau}$.*

Proof. Suppose by contradiction that $\psi_j \leq \frac{1}{\tau}$ in the IC equilibrium. Let $q_{jl}^{IC}$ denote the IC equilibrium allocation from (D.7). Because $\psi_j \leq \frac{1}{\tau}$, we have that $q_{jl}^{IC} \geq q_{jl}^{IR}$, where the latter is defined in (D.10). From the definition of $\psi_j$, (D.8), this implies that $\psi_j < \frac{1}{\tau}$ for all $q_{jl} \leq q_{jl}^{IR}$. Now let's consider whether the IR allocation satisfies the IC constraint (D.13). At $q_{jl} = 0$, that equation holds with equality (trivially, $0 = 0$). The derivative of the LHS of (D.13) with respect to $q_{jl}$ is

$$\frac{\partial LHS(D.13)}{\partial q_{jl}} = \frac{\sigma-1}{\sigma}\left(q_{-jl}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}} - \tau\frac{\sigma-1}{\sigma}\left(q_{-jh}^{\frac{\epsilon-1}{\epsilon}} + q_{jl}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}-1} q_{jl}^{-\frac{1}{\epsilon}}.$$

For this to be positive, we need

$$\psi_j(q_{jl}) \leq \frac{1}{\tau}.$$

As we showed above, this is true for all $q_{jl} \leq q_{jl}^{IR}$. Hence, the IC is satisfied in the IR allocation. Contradiction.

■

# E Nonlinear costs

In this section, we consider a modification to the baseline model where firms use flexible cost functions. Instead of assuming firms face a linear production cost, we only assume that total production cost is a function of total production.

**Setup.** Let the production cost of firm $j$ be denoted by $C_j(q_j)$ where $q_j = \pi q_{\tau j} + (1-\pi)q_{1j}$ is the total production of firm $j$. Then, the firm's problem is given by

$$\begin{aligned}
\max_{\{q_{1j},q_{\tau j},p_{1j},p_{\tau j}\}} \quad & \pi q_{\tau j}p_{\tau j} + (1-\pi)q_{1j}p_{1j} - C_j(\pi q_{\tau j} + (1-\pi)q_{1j}) && \text{(E.1)}\\
\text{s.t} \quad & u(q_{1j}) - \frac{p_{1j}q_{1j}}{P} = 0 && [IR_1]\\
& \tau u(q_{\tau j}) - \frac{p_{\tau j}q_{\tau j}}{P} = \tau u(q_{1j}) - \frac{p_{1j}q_{1j}}{P} && [IC_\tau]
\end{aligned}$$

Denote by $c_j(q_j) \equiv C_j'(q_j)$ the marginal cost of production of firm $j$. Then, the optimal quantity of firm $j$ is given by

$$\tau\, u'(q_{\tau j}) = \frac{c_j(q_j)}{P}, \tag{E.2}$$

$$u'(q_{1j}) = \frac{1-\pi}{1-\tau\pi}\,\frac{c_j(q_j)}{P}. \tag{E.3}$$

These optimality conditions are very similar to our benchmark model equations (2.12)–(2.13)

**Efficient allocation.** The planner's problem, who is constrained to use the same level of aggregate labor, is given by[64]

$$\max_{\{q_{1j},q_{\tau j}\}} \quad \int_j [\pi u(q_{\tau j}) + (1-\pi)u(q_{1j})]\, dj \tag{E.4}$$

$$\text{s.t.} \quad \int_j C_j\left(\pi q_{\tau j} + (1-\pi)q_{1j}\right) = \bar{L}. \tag{E.5}$$

Let $P^{AE}$ denote the Lagrange multiplier on the planner's resource constraint. Then, the efficient allocation quantities satisfy

$$u'(q_{\tau j}^{\mathrm{AE}}) = \frac{c_j(q_j^{AE})}{\tau}\,\frac{1}{P^{\mathrm{AE}}}, \tag{E.6}$$

$$u'(q_{1j}^{\mathrm{AE}}) = c_j(q_j^{AE})\,\frac{1}{P^{\mathrm{AE}}}, \tag{E.7}$$

for all $j$, where $q_j^{AE} \equiv \pi q_{\tau j}^{AE} + (1-\pi)q_{1j}^{AE}$ is the quantity produced by firm $j$ in the efficient allocation.

[64]In the planner's problem we omit the IC constraints as those will not bind in equilibrium. The proof of this argument mirrors the proof of Proposition 1

## E.1 Propositions and Proofs

In this section, we prove that the two main results of our benchmark model still hold: firm-level output and employment are identical to the efficient allocation, but high-taste consumers are sold too much, and low-taste consumers too little of each good.

Proof of Proposition 4 with nonlinear production costs.

Equations (E.2–E.3), together with the concavity of $u(\cdot)$, imply that the production of all firms is increasing in the aggregate price index $P$. Therefore, there is a unique level of the aggregate price index that clears the labor market.

Let $\widetilde{P}_j$ be the aggregate price index such that the firm-level production of a firm with production cost function $C_j$ in equilibrium is identical to its overall production in the efficient allocation: $(1-\pi)\left[q_{1j}^{\mathrm{AE}} - q_{1j}\right] - \pi\left[q_{\tau j} - q_{\tau j}^{\mathrm{AE}}\right] = 0$. Using (E.3–E.2), this can be written as:

$$(1-\pi)\left[(u')^{-1}\left(\frac{c_j(q_j^{AE})}{P^{AE}}\right) - (u')^{-1}\left(\frac{1-\pi}{1-\tau\pi}\frac{c_j(q_j^{AE})}{\widetilde{P}_j}\right)\right] - \pi\left[(u')^{-1}\left(\frac{c_j(q_j^{AE})}{\tau\widetilde{P}_j}\right) - (u')^{-1}\left(\frac{c_j(q_j^{AE})}{\tau P^{AE}}\right)\right] = 0. \tag{E.8}$$

Assumption 1 implies that $\partial\log(q_{\tau j} - q_{\tau j}^{AE})/\partial\log(c_j(q_j^{AE})) = \eta$. This follows from Equation (3.2), when relabeling $x = c_j(q_j^{AE})/(\tau\tilde{P}_j)$ and $\tau = \tilde{P}_j/P^{AE}$. Similarly, $\partial\log(q_{1j}^{AE} - q_{1j})/\partial\log(c_j(q_j^{AE})) = \eta$.

Now consider a firm with $c_k(q_k^{AE}) = (1+\Delta)c_j(q_j^{AE})$. Using Assumption 1, we have that

$$\begin{aligned}\pi\Big(q_{\tau,k}(\tilde{P}_j) - q_{\tau,k}^{AE}\Big) \qquad\qquad & - (1-\pi)\left(q_{1,k}^{AE} - q_{1,k}(\tilde{P}_j)\right) = \\ \pi(1+\Delta)^{\eta}\left(q_{\tau,j}(\tilde{P}_j) - q_{\tau,j}^{AE}\right) & - (1-\pi)(1+\Delta)^{\eta}\left(q_{1,j}^{AE} - q_{1,j}(\tilde{P}_j)\right) = 0.\end{aligned}$$

Since there is a unique level of the aggregate price index such that the labor market clears, it must be that $P = \widetilde{P}_j$. Hence, the equilibrium firm-level production and employment for all firms is identical to the ones in the efficient allocation.

■

Proof of Proposition 3 with nonlinear production costs and CED preferences. From the Proposition above, we know that $q_j^{AE} = q_j$. That is, the firm-level quantity sold in the decentralized equilibrium is equal to the efficient allocation level. From equations (E.2–E.3) and (E.6), together with noting that the marginal cost of firm $j$ is the same across the two equilibria, we have that:

$$\frac{u'(q_{\tau j})}{u'\left(q_{\tau j}^{AE}\right)} = \frac{P^{AE}}{P}, \tag{E.9}$$

$$\frac{u'(q_{1j})}{u'\left(q_{1j}^{AE}\right)} = \frac{1-\pi}{1-\tau\pi}\frac{P^{AE}}{P}. \tag{E.10}$$

The equations above, together with the fact that $u'(q)$ is decreasing in $q$ imply that one of three cases must hold: (i) if $\frac{P}{P^{AE}} > 1$ then $q_{\tau j} > q_{\tau j}^{\mathrm{AE}}$ and $q_{1j} > q_{1j}^{\mathrm{AE}}$ for all $j$, (ii) if $\frac{P}{P^{AE}} \in \left(\frac{1-\tau\pi}{1-\pi}, 1\right)$ then

$q_{\tau j} > q_{\tau j}^{\mathrm{AE}}$ and $q_{1j} < q_{1j}^{\mathrm{AE}}$ for all $j$, and (iii) if $\frac{P}{P^{AE}} < \frac{1-\tau\pi}{1-\pi}$ then $q_{\tau j} < q_{\tau j}^{\mathrm{AE}}$ and $q_{1j} < q_{1j}^{\mathrm{AE}}$ for all $j$.

Aggregate labor market clearing implies that

$$\int_0^1 c_j \left(\pi q_{\tau j} + (1-\pi) q_{1j}\right) dj = \int_0^1 c_j \left(\pi q_{\tau j}^{\mathrm{AE}} + (1-\pi) q_{1j}^{\mathrm{AE}}\right) dj,$$

so that neither option (i) nor option (iii) are consistent with equilibrium. Therefore, it must be that $\frac{P}{P^{AE}} \in \left(\frac{1-\tau\pi}{1-\pi}, 1\right)$, so that $q_{\tau j} > q_{\tau j}^{\mathrm{AE}}$ and $q_{1j} < q_{1j}^{\mathrm{AE}}$ for all $j$. ■