CESifo AREA CONFERENCES 2020

Global Economy
Online, 17–18 September 2020

Input-Output Linkages and Monopolistic Competition: Input Distortion and Optimal Policies

Benjamin Jung and Wilhelm Kohler

CESifo

Input-Output Linkages and Monopolistic Competition: Input Distortion and Optimal Policies*

Benjamin Jung[†] and Wilhelm Kohler[‡]

February 29, 2020

Abstract

We show that the combination of monopolistic competition and input-output linkages generate what we call an input distortion. The distortion arises because material input prices involve a markup over the social opportunity cost. This has so far escaped attention in the literature addressing efficiency of monopolistic competition equilibria. Using a stylized single sector model, we provide a full description of the social optimum for an economy featuring an input-output linkage in the presence of monopolistic competition. Using this as a benchmark, we describe the allocational inefficiency of a decentralized market equilibrium as well as first-best policies to achieve efficiency. In an integrated world equilibrium, a material input subsidy and an output subsidy turn out to be perfect substitutes. A wage tax is unable to serve in offsetting the input distortion. In a cooperative policy setting with two countries, an input subsidy is a second-best policy to address the input distortion.

JEL-Classification: F12, F13, D57, D61, H21
Keywords: input-output linkages, monopolistic competition, international trade, allocational inefficiency, optimal policy

*We are grateful to Carsten Eckel, Hartmut Egger, Peter Egger, Gabriel Felbermayr, and Jens Südekum as well as participants of the Göttinger Workshop "Internationale Wirtschaftsbeziehungen" for helpful comments and discussions. Benjamin Jung gratefully acknowledges support by the Faculty of Economics and Social Sciences at the University of Hohenheim within its research focus "Inequality and Economic Policy Analysis" (INEPA). Wilhelm Kohler is grateful for financial support received from Deutsche Forschungsgemeinschaft (DFG) under grant no. KO 1393/2-1.

[†]University of Hohenheim, 70593 Stuttgart, Germany; CESifo; IAW; jung.benjamin@uni-hohenheim.de.

[‡]University of Tübingen; IAW; CESifo, GEP; wilhelm.kohler@uni-tuebingen.de

1 Introduction

The misallocation of resources plays a major role in explaining cross-country differences in per-capita incomes (Jones, 2013). Production inefficiencies are also an important justification of government intervention. An important source of production inefficiencies is imperfect competition, or market power. Arguably, production inefficiencies due to market power are best dealt with through competition policy, but very often this doesn't lead very far. For instance, if market power derives from product differentiation, then it is not only a source of production inefficiency, but also a 'by-product' of an important source of welfare, viz. product variety. In this case, avoiding market power is both, practically difficult and potentially harmful. Yet market power *may* still cause production efficiencies, begging the question of what governments might do in order to improve upon the market outcome. Very often, the expected answer is taxes and/or subsidies.

However, given the nexus of market power and product differentiation, whether or not there is a true distortion calling for government intervention is not entirely clear. For instance, it is well known from recent trade literature that under certain conditions a zero profit equilibrium under monopolistic competition is efficient, although as we shall see below these conditions are rather restrictive. In this paper, we demonstrate that the presence of input-output linkages renders a monopolistic-competition-equilibrium inefficient even under conditions where existing literature would diagnose efficiency. By input-output linkage, in general terms what we mean is the production of commodities by means of commodities, goods serving as inputs in *their own* production. In the presence of monopolistic competition they generate a specific form of production inefficiency that, to the best of our knowledge, has so far gone unnoticed in the literature.

Our production inefficiency relates to what is sometimes called the *roundaboutness of production*. The underlying assumption is that production technology allows for substitution between the primary input (labor) and the material input – the good produced serving as an input in its own production. The latter involves forgone consumption and, thus, a roundabout way of turning labor into consumption. Naturally, the primary input is essential in production, and the material input is an imperfect substitute for labor. Under these assumptions there exists an optimal degree of roundaboutness in production, and our key proposition in this paper is that if production involves monopolistic competition, as commonly used in much of modern trade or macroeconomic literature, then a decentralized market equilibrium features a specific form of production inefficiency: There is too much direct labor use and not enough use of the produced input. In other words, there is suboptimally low degree of roundaboutness in production.[1] Throughout this paper, we shall refer to this as the *input distortion*.

This type of inefficiency is absent if the use of intermediates simply reflects fragmentation of the production process into several stages. On each stage, the cost of the intermediate might be distorted if the intermediate is offered with market power, and with multiple stages there might be multiple margins. As a consequence, the final good becomes available at a distorted price and will give rise to misallocation of resources across sectors, unless the price distortion is uniform across sectors. Naturally, no such misallocation can arise in a single-sector-world, provided production is *recursive*, meaning that no good is ever used as an input in its own production. In contrast, a true input-output linkage makes the production relationship non-recursive, in which case markup-pricing on intermediate inputs gives rise to an entirely different form of misallocation, viz. the aforementioned input distortion.

We discuss this production inefficiency against the backdrop of the past three decades in which models of monopolistic competition have become increasingly important in several branches of economics. Krugman (1979,1980) and Ethier (1982) have introduced such models into trade theory where they have quickly acquired workhorse status, but the monopolistic competition paradigm is widely used in other areas of economics as well, such as new Keynesian macroeconomics (Blanchard and Kiyotaki, 1987), or endogenous growth (Romer,

[1] There is a long tradition in the literature, dating back to von Böhm-Bawerk, of discussing this idea of roundaboutness in the context of growth theory, where roundaboutness involves the lapse of time, which becomes key of individuals have time preference. See Hennings, K.H. (1987), “Capital as a Factor of Production,” in: Palgrave Macmillan (eds), The New Palgrave Dictionary of Economics, Palgrave Macmillan, London; and Hennings, K.H. (1987), Roundabout Methods of Production, in: Palgrave Macmillan (eds), The New Palgrave Dictionary of Economics, Palgrave Macmillan, London. This paper, however, is intended as a contribution, not to this literature, but - first and foremost - to modern trade literature where input-output linkages have played an important role which is more or less orthogonal to its role in growth theory.

1990). More recently, Melitz (2003) has introduced firm heterogeneity into a model of monopolistic competition, thus increasing the degree of realism and providing further impetus to the widespread use of this model.[2] A cornerstone of established knowledge relating to monopolistic competition is that in a *single sector* world with a *constant elasticity of substitution* (CES) in demand, the decentralized laissez faire equilibrium is efficient. The reason is that with standard CES preferences, "consumer surplus" and "profit destruction" distortions exactly cancel out, as pointed out by Baldwin (2005). Moreover, since all firms charge the same markup consumers' cross-firm spending decisions are not distorted. Importantly, in Melitz-type models with firm heterogeneity the efficiency-result also applies to the equilibrium mass of firm entry into the market and to the selection of firms taking up production (Dhingra and Morrow, 2019).

However, efficiency is lost if the degree of external economies from product variety differs from the degree implicit in standard CES models, or if the elasticity of substitution in demand is variable; see Benassy (1996,1998) and Dhingra and Morrow (2019). Production efficiency is also lost in multi-sector models with CES preferences, as indicated above. Generally, industries with above average markups of prices over marginal cost will produce in less than optimal amounts; see Epifani and Gancia (2011) for a framework with constant markups and Behrens et al. (2018) for a model that features variable markups. Caliendo et al. (2017) explore the implications of the entry distortion in multi-sector models for optimal trade policy.

Input-output linkages in monopolistic competition environments were first analyzed by Krugman and Venables (1995) and Venables (1996). The focus there, as in the entire literature on new economic geography, lies on input-output linkages as a mechanism responsible for the emergence of agglomeration equilibria. The input distortion highlighted in this paper goes unnoticed. Input-output linkages have played an important role also in large-scale computable general equilibrium models, some of which also feature monopolistic competition (e.g., Balistreri et al., 2011), but again, this literature has not identified or analyzed the input distortion. The same holds true for the recent literature on "quantitative trade models"

[2]Witness the first four chapters in the most recent volume no 4 of the Handbook of International Economics (2014).

surveyed by Costinot and Rodríguez-Clare (2014).[3] A paper that comes relatively close to the present one is Caliendo et al. (2017). They model input-output linkages in a way completely analogous to this paper, and they focus on the implications of entry distortions in the presence of firm heterogeneity. We shall return to the similarities and differences between this paper and ours in more detail below.[4]

We analyze the input distortion using a formal model with monopolistic competition and input-output linkages as familiar from the trade literature. However, to bring our novel point into sharp focus, we choose a model structure that rules out all of the distortions so far identifies in the literature. In particular, we assume a single sector and we model monopolistic competition employing the CES-version of the variety effect in production, as in Ethier (1982). We assume two economies (home and foreign), assembling an aggregate good from their own varieties, respectively, according to a CES-technology, where the elasticities of substitution are allowed to differ across countries. Each country has a further layer of assembly where it uses its own as well as the other country's aggregate good, again based on a CES technology with a common elasticity of substitution between the two countries' goods (Armington assumption). This is where goods trade enters our stage. On both layers, assembly is governed by perfect competition, and within each country assembly generates a final good which is available for the country's own consumption, but which is also used as an input into the country's production of varieties. Importantly, production of varieties is governed by monopolistic competition, leading to markup-prices for varieties used in assembly. Ultimately, these markups lead to a price of the final good which is above its opportunity cost. Given a single sector and inelastic labor supply, this does not distort consumption, but the input-output linkage implies that it distorts the variety producers' choice of the input mix between the produced input and the primary input – the input distortion.

[3]To be clear, this does not, per se, invalidate the results that such models generate regarding the positive and normative consequences of specific trade liberalization scenarios. But awareness of the above type of production inefficiency and the corresponding mechanisms should contribute to a full understanding of the numerical results obtained. Of course, models relying on perfect competition, such as Caliendo and Parro (2015), based on Eaton and Kortum (2002), generally run the risk of ignoring important mechanisms, including in particular the input distortion highlighted in this paper.

[4]Blanchard et al. (2017) explore optimal final good tariffs in the presence of global value chains, but use a different modeling approach and do not pay attention to the underlying distortions.

In the first step of our analysis, we focus on the integrated world economy. To establish a reference case, we provide a complete solution of the social optimum, to be compared with the decentralized market equilibrium in the presence of different types of policy interventions. We shall identify different types of subsidization that may serve as a first-best policy to correct the input distortion. We also explore the implications of several types of extensions regarding the specific way that the input-output linkage appears in the economy. For instance, we allow for the input distortion to also affect the fixed cost of variety producers. In the polar case where input-output linkages only occur in the fixed but not in the variable input, material input is efficient, but the labor use is inefficiently large, again resulting in too much direct labor use, relative to indirect labor use. With inefficiently large labor input, the number of firms is inefficiently low, which constitutes an additional channel through which the input distortion affects total factor productivity and consumption (welfare). As regards policy, we also explore situations in which subsidies on domestic intermediate inputs are not available.[5] In this case, cooperative subsidies on imports serve as a second-best instrument to address the input distortion.[6] And finally, we discuss why a wage tax will not be a suitable policy measure to correct the input distortion, although at first sight it may seem like an offsetting distortion. A corollary to this is that wage distortions generated by wage formation under different types of labor market imperfections will likewise not serve as a distortion that potentially offsets our input distortion.

The second step of our analysis, we generalize our main results to settings that feature country borders regarding labor endowments and costly trade in intermediates. We also discuss trade policies in the form of an import tax or subsidy. An import subsidy turns out to be a second-best instrument to address the input distortion. This finding implies that in a non-cooperative environment, the effect of trade policy on the roundaboutness-margin are to be weighed against the familiar terms-of-trade effects of trade policy.[7]

[5] The Agreement on Subsidies and Countervailing Measures disciplines the use of production subsidies.

[6] In our setting, trade policy cannot be conditioned on the type of use of the imported good. While such discriminatory instruments are not strictly ruled out, they are well known to be difficult and costly to maintain, since they are liable to cause evasion.

[7] For optimal non-cooperative trade policy in monopolistic competition settings without input-output link-

The remainder of the paper is structured as follows. Section 2 develops a model with monopolistic competition and an input-output linkage. The integrated world equilibrium and the setting with country borders and trade costs are analyzed in sections 3 and 4, respectively. The final section concludes.

2 Input-output linkages in a two-country framework

We use a highly stylized model adopting features familiar from the literature in all aspects except input-output linkages. This allows us to sharply focus on our main point. We assume two countries, home (h) and foreign (f), producing large numbers N_h and N_f, respectively, of differentiated goods, which symmetrically enter production of a composite (or aggregate) good according to a CES production function with an (Armington) elasticity of substitution $\sigma = 1/(1-\rho) > 1$:

$$Y_i = \left(N_i q_i^{\rho_i}\right)^{1/\rho_i} = N_i^{\nu_i+1} q_i, \quad i \in \{h, f\} \tag{1}$$

$$Y_i = Y_{ii} + \delta Y_{ij}, \quad i, j \in \{h, f\},\ i \neq j, \tag{2}$$

$$Q_i = \left(Y_{ii}^{\rho} + Y_{ji}^{\rho}\right)^{1/\rho}, \quad i, j \in \{h, f\},\ i \neq j. \tag{3}$$

In these equations, Q_i is the quantity of the final good produced and consumed in country i, while Y_i is the quantity of a country-specific aggregate good, assembled from N_i varieties produced in country i, and used in amount Y_{ii} for country i's own final good assembly. The remainder is available for the final good assembly in country j, albeit subject to an iceberg-type trade cost $\delta > 1$. The parameter $\rho_i \in (0, 1)$ measures the degree of substitution between different varieties produced in country i, related to the elasticity of substitution $\sigma_i > 1$ according to $\rho_i = (\sigma_i - 1)/\sigma_i$. It is also inversely related to the strength of economies from enhanced variety, measured by $\nu_i = 1/(\sigma_i - 1) > 1$.

---

ages, see Gros (1987), Demidova and Rodríguez-Clare (2009), Felbermayr et al. (2013), and Costinot et al. (2016).

Differentiated goods are produced using a Cobb-Douglas production technology:

$$q_i = \left(\frac{\ell_i}{\gamma}\right)^{\gamma} \left(\frac{m_i}{1-\gamma}\right)^{1-\gamma}, \quad i \in \{h, f\}, \tag{4}$$

where ℓ_i denotes the quantity of labor input and m_i denotes the quantity of the aggregate good used as an input in variety production, henceforth called the material input. The parameter $\gamma \in (0, 1]$, assumed the same for both countries, is the elasticity of output with respect to labor. The standard model of monopolistic competition thus emerges if $\gamma = 1$. Production of each differentiated good is subject to increasing returns to scale, which arise from the presence of a fixed research and development (R&D) cost that a firm has to incur in order to enter the market. In our baseline model, this fixed cost is incurred in terms of labor, and denoted by f, again assumed the same for both countries. In a model extension, we allow for intermediate inputs also in the fixed cost part of the technology. For the sake of a clear and simple demonstration of the input distortion, we abstract from Melitz-type selection effects driven by firm heterogeneity. We shall see, however, that selection effects would not alter the main results of our paper.

Aggregate output Q_i is nontradable and may be used for consumption C and as an intermediate input in production of varieties. Total intermediate input use is given by

$$M_i = N_i m_i, \quad i \in \{h, f\}. \tag{5}$$

We assume a linear utility function whence our welfare measure reads as

$$U_i(C_i) = C_i = Q_i - M_i > 0, \quad i \in \{h, f\} \tag{6}$$

Note that, given the input-output linkage, we have to observe a viability constraint as evidenced by the inequality sign in the above expression.[8] We assume a given labor endowment

[8] In input-output analysis, this viability condition is known as the *Hawkins-Simon condition*. Writing the vector of final consumption quantities as $\mathbf{c}$ and assuming a Leontief-technology with input coefficients collected in a Matrix $\mathbf{A}$, the Hawkins-Simon condition imposes a restriction on $\mathbf{A}$ guaranteeing that there exists a vector $\mathbf{x}$, such that $(\mathbf{I} - \mathbf{A})\mathbf{x} \geq 0$. In our context, Q is the equivalent of $\mathbf{I}\mathbf{x}$, aggregate (gross) output, while $\mathbf{M}$ is the

equal to L, so that the resource constraints of the two economies require

$$L_i \geq N_i\left(\ell_i + f\right), \quad i \in \{h, f\} \tag{7}$$

Throughout the paper we assume that this condition is satisfied with equality.

3 Integrated world equilibrium

In order to obtain a crystal-clear view on the inefficiency of the laissez-faire market equilibrium, we first characterize the integrated world economy. This economy, recognized by absence of subscripts, has a labor endowment equal to $L = L_h + L_f$, and it has no real trade cost, $\delta = 1$. To simplify things further, we also assume $\rho_h = \rho_f = \rho$ as well as $L_h = L_f$, whence we may set $N_h = N_f$ and $q_h = q_f = q$. Due to complete symmetry the production relationships (1) through (3) simplify to

$$Q = N^{\nu+1}q, \tag{8}$$

where $N = N_h + N_f$. The endowment constraint now reads as $L = N\left(\ell + f\right)$, and world welfare emerges as $C = Q - M$, where $M = Nm$.

3.1 Social optimum

Society faces two trade-offs. The first relates to the roundaboutness of labor use. Increasing m/ℓ economizes on labor in production of differentiated goods intermediates, but it also reduces the amount of output available for consumption. The second is the trade-off between variety and efficiency: Increasing the number of differentiated varieties increases output even for a constant intermediate input use, but at the same time it reduces output available for

equivalent of $\mathbf{Ax}$, the inputs needed to generate $\mathbf{x}$. Obviously, in strict form the Hawkins-Simon condition is equivalent to the inequality in (6) above. In input-output analysis, the coefficients $\mathbf{A}$ are given exogenously. In our single sector case, the intermediate input intensity of production is chosen endogenously, and the equivalent of the Hawkins-Simon condition will be seen to hold from the first-order condition on intermediate input use.

consumption on account of the fixed cost. The social optimum lies in the solution of the following problem:

$$\max_{\ell,m,N} C = Q - M \quad \text{s.t.} \quad L = N\left(\ell + f\right),$$

where $C = Q - M$, where $M = Nm$. The corresponding Lagrangian is

$$\pounds = Q - M - \lambda\left[N\left(l + f\right) - L\right].$$

The marginal productivity of labor and the material input, respectively, in variety production are equal to $\gamma q/\ell$ and $(1-\gamma)q/m$. Using * to indicate optimal values, the first-order conditions for ℓ^* and m^* are

$$\left(N^*\right)^{\nu+1}\gamma q^*/\ell^* = \lambda^* N^* \quad \Rightarrow \quad \ell^* = \gamma\frac{Q^*}{\lambda^* N^*} \tag{9}$$

$$\text{and} \quad \left(N^*\right)^{\nu+1}(1-\gamma)q^*/m^* = N^* \quad \Rightarrow \quad m^* = (1-\gamma)\frac{Q^*}{N^*} \tag{10}$$

In (9), λ^* denotes the shadow value of labor. Note that the conditions on the two types of inputs are similar, but that optimal material input level m^* does not depend on the shadow value of labor. The reason is that the opportunity cost of intermediate input use in production is foregone consumption *of the same good*. Note also that equation (10) implies $M^* = N^*m^* = (1-\gamma)Q^*$. Since $\nu > 0$ and $\lambda^* > 0$, the condition (11) also ensures viability, meaning that $Q^* > M^*$; see footnote 8 above.

The first-order condition on N^* is $(\nu+1)\left(N^*\right)^{\nu} - m^*\lambda^*(\ell^* + f)$, which may be written as

$$\left(\nu + 1\right)Q^* - M^* = \lambda^* L. \tag{11}$$

This condition simply states that the value deriving from additional *net* output obtained from a marginal increase of N must be equal to the marginal cost of N, which is equal to $\lambda^* L/N$. Note that this cost includes variable and fixed labor input into production. Finally, the derivative with respect to the Lagrange multiplier λ yields the resource constraint (7) in its binding form.

Using these first-order conditions, we may now solve for the optimal levels of ℓ^*, m^*, N^*, and λ^*. The optimal labor intensity of variety production is given by

$$\frac{\ell^*}{m^*} = \frac{\gamma}{1-\gamma} \Big/ \lambda^*. \tag{12}$$

Given $M^* = (1-\gamma)Q^*$, the condition (11) may be rewritten as

$$(\nu + \gamma)Q^* = \lambda^* N^* (\ell^* + f). \tag{13}$$

Using the first equation in (9) to replace Q^*, we obtain the following variable labor use per firm:

$$\ell^* = \frac{\gamma}{\nu} f. \tag{14}$$

Intuitively, the variable labor use relative to labor required for the fixed cost is increasing γ and falling in ν, the degree of economies from product variety.

Solving the resource constraint for N^* and replacing for ℓ^*, we obtain

$$N^* = \frac{1}{\gamma/\nu + 1} \frac{L}{f}. \tag{15}$$

As is common in the literature, we ignore the integer constraint on N, but allowing a value of N^* below unity does not make sense. Indeed, when turning to the decentralized equilibrium below we assume the number of firms to be large. Hence, we implicitly assume $L/f > \gamma/\nu + 1 = \gamma(\sigma - 1) + 1$. Compared to the standard model without input-output linkages ($\gamma = 1$), we recognize that input-output linkages increase the efficient number of firms, which implies a lower ratio of variable to fixed labor use per variety produced. Thus, input-output linkages strengthen (weaken) the positive (negative) relationship between economies from product variety on the one hand (captured by $\nu > 1$), and the optimum number of varieties (firm size) on the other. Intuitively, since production of any variety draws on intermediate inputs as well as labor, a given "endowment ratio" L/f allows for a larger number of production facilities (varieties).

Using the first-order condition on m^*, replacing $Q^* = N^{*\nu+1} q^*$, using the production function for q^*, and inserting the solution for N^* from above, we arrive at

$$m^* = \frac{(1-\gamma)\, f}{\nu} \left(\frac{1}{\gamma/\nu + 1} \right)^{\frac{v}{\gamma}} \left(\frac{L}{f} \right)^{\frac{\nu}{\gamma}} . \tag{16}$$

The term $(1-\gamma) f/\nu$ in an intuitive way mirrors (14) for optimal variable labor use. In addition, comparing the second term in (16) to (15), we recognize that the optimal intermediate input use per firm is also increasing in the number of firms. This essentially reflects the resource constraint of the economy. A higher number of firms implies a higher use of labor for fixed cost.

Taking the above solutions, we may write

$$\frac{m^*}{\ell^*} = \frac{1-\gamma}{\gamma} \left(N^* \right)^{\nu/\gamma} . \tag{17}$$

Equating this to the right-hand side of (12) finally allows us to determine the shadow value of labor as

$$\lambda^* = \left(\frac{1}{\gamma/\nu + 1} \right)^{\nu/\gamma} \left(\frac{L}{f} \right)^{\nu/\gamma} . \tag{18}$$

Remember that this is expressed in terms of consumption. Intuitively, it increases with the ratio of L/f, due to aggregate economies of scale from the number of varieties of inputs in final good assembly. The parameter restriction $L/f > \gamma/\nu + 1$ now implies that $\lambda^* > 1$.

It is instructive to investigate the degree of roundaboutness in the social optimum. We use $A_\ell^* := \ell^* N^*/Q^*$ to denote the direct labor input coefficient in the variable cost part of production and accordingly for the direct material input coefficient, $A_m^* := m^* N^*/Q^* = 1 - \gamma$. Adding indirect material input use, the total material input coefficient may be written as $A_m^* \left(1 - A_m^*\right)^{-1} = (1-\gamma)/\gamma$. The labor embodied in total material input is $A_\ell^*(1-\gamma)/\gamma$, hence the ratio $(1-\gamma)/\gamma$ gives the ratio between direct and indirect labor use, looking only at variable inputs. From (14), the fixed labor input per unit of the aggregate good is $f N^*/Q^* = (\nu/\gamma)\, \ell^* N^*/Q^* = (\nu/\gamma)\, A_\ell^*$, which is all direct labor use. A useful measure of roundaboutness is the ratio of indirect labor use, $A_\ell^*(1-\gamma)/\gamma$, relative to total labor use $A_\ell^*(1+\nu/\gamma+(1-\gamma)/\gamma)$.

Denoting this measure by R, we have

$$R^* := \frac{1-\gamma}{1+\nu}, \tag{19}$$

which lies between zero and one.[9] Remember that $\nu > 1$ measures the degree of economies from product variety. A high value of ν means that it is economical to have many firms. But with many firms more of the overall labor endowment must be devoted to the fixed input, which does not involve any roundaboutness but only direct labor use. Hence R^* decreases in ν.

Finally, the level of welfare (consumption) in the social optimum may be derived as follows. First, we have $C^* = \gamma Q^* = \gamma (N^*)^{\nu+1} q^*$. Since $m^* = \ell^* \lambda^* (1-\gamma)/\gamma$, we may write $q^* = (\ell^*/\gamma)(\lambda^*)^{1-\gamma}$, whence $C^* = \gamma Q^* = \ell^* (N^*)^{\nu+1} (\lambda^*)^{1-\gamma}$. Using $\lambda^* = (N^*)^{\nu/\gamma}$ from above and substituting $\ell^* = \gamma f/\nu$, we obtain

$$C^* = \frac{\gamma f}{\nu} \left(\frac{1}{\gamma/\nu + 1} \frac{L}{f} \right)^{(\gamma+\nu)/\gamma} = \frac{\gamma}{\nu} \left(\gamma/\nu + 1\right)^{-(\gamma+\nu)/\gamma} \left(\frac{L}{f} \right)^{\frac{\nu}{\gamma}} L. \tag{20}$$

Equivalently, welfare may be written as

$$\frac{C^*}{L} = \frac{\gamma}{\gamma+\nu} \lambda^* \tag{21}$$

Consumption per capita is lower than the marginal shadow value of labor, as expected in a situation with (exogenously) increasing returns to scale ($\nu > 1$). This is reinforced by the input output linkage ($\gamma < 1$). For later reference, we also note that total factor productivity (TFP) in the social optimum is

$$\text{TFP} = \frac{Q^*}{L} = \frac{1}{\gamma+\nu} \lambda^* \tag{22}$$

which follows from $C^* = \gamma Q^*$.

The following proposition summarizes the role the input-output linkages play in the so-

[9] Alternatively, we might measure roundaboutness as the ratio between indirect and direct labor use, inclusive of the fixed labor input: $[(1-\gamma)/\gamma]/\,(1+\nu/\gamma) = (1-\gamma)/(1+\nu)$, which lies between zero and infinity.

cial optimum.

Proposition 1 ***(a)*** *For a given ratio L/f, consumption per capita is a constant. Any increase in L/f leads to an over-proportional increase in per capita consumption ($\nu/\gamma > 1$).* ***(b)*** *An increase in L/f also raises the shadow value of labor; in elasticity terms, this relationship is reinforced by the input-output linkage.* ***(c)*** *An increase in L/f lowers the labor intensity ℓ^*/m^*, but the optimal degree of roundaboutness in production, measured by the ratio of indirect to total use of labor, is given parametrically by $\frac{1-\gamma}{1+\nu}$.* ***(d)*** *The level of consumption per capita is below the marginal shadow value of labor, and this discrepancy is reinforced by the input-output linkage.*

Proof. Part **(a)** follows from (20). Part (**b**) follows from (18). Part (**c**) follows from (17) and (18) as well as 19). And (**d**) follows writing (20 as $C^* = \frac{\gamma}{\gamma+\nu}\lambda^* L$. ∎

In the present setup, γ is a primitive of the technology and not subject to policy influence. It is nevertheless instructive to explore how a change in technology, say a strengthening of the input-output linkage, represented by a reduction in γ, affects maximum consumption per capita. Equation (20) seems to suggest that the relationship between the strength of input-output linkages and the maximum level of consumption per capita, C^*/L, is ambiguous. A reduction in γ has two opposing effects. First, it lowers the share of aggregate output available for consumption, and secondly, it raises aggregate output. The first effect follows directly from $C^* = (1 - M^*/Q^*)Q^* = \gamma Q^*$. The second effect follows from

$$Q^* = \frac{f}{\nu}\left(N^*\right)^{\nu/\gamma+1}, \tag{23}$$

which in turn follows from equations (14) through (16) above. It turns out that under the above mentioned parameter restriction the first effect always dominates:

Proposition 2 *In the social optimum, a reduction in γ (which implies a higher roundaboutness of labor use) leads to an increase in the level of consumption per capita.*

Proof. Taking logs in (20) and differentiating with respect to γ, we obtain

$$\frac{\partial \ln C^*}{\partial \gamma} = \frac{\nu}{\gamma^2} \left[\ln \left(\frac{\gamma}{\nu} + 1 \right) - \ln \left(\frac{L}{f} \right) \right].$$

Given the parameter restriction discussed in connection with N^* subsequent to equation (15), it follows that $\frac{\partial C^*}{\partial \gamma} = C^* \frac{\partial \ln C^*}{\partial \gamma} < 0$. ■

Why should a higher intermediate input intensity of production (lower γ) lead to a higher level of aggregate output Q^*, as evidenced by equations (23) and (15)? Intuitively, a lower γ saves on direct labor use in production of varieties, thus freeing up labor for fixed input use. Due to economies from enhanced variety, the indirect labor requirement for production of the additional intermediate inputs (in line with a lower γ) is lower than the incipient reduction in the direct use of labor. If the initial number of firms were equal to one, then the percentage increase in aggregate output would be exactly equal to the percentage reduction in direct labor use. But if $N^* > 1$, then the percentage increase in aggregate output is less than the percentage savings on direct labor use per initial variety produced. By implication, a higher degree of roundaboutness in production is a source of welfare increase.

3.2 Decentralized equilibrium with policy intervention

We now characterize a decentralized market equilibrium, in which producers of varieties behave under monopolistic competition, while assembly of the final good is governed by perfect competition. Writing $\tilde{p}$ for the "demand-price" for a differentiated variety faced by final good producers, the minimum unit-cost function dual to (8) is $N^{-\upsilon}\tilde{p}$. Using $\tilde{P}$ to denote the price of the final good, zero profits in final goods production implies

$$\tilde{P} = N^{-\upsilon}\tilde{p}. \tag{24}$$

Unit-demand for a variety follows as

$$q = \left(\frac{\tilde{p}}{\tilde{P}} \right)^{-\sigma}. \tag{25}$$

Given what we said in the introduction about the input distortion, it is useful to introduce a policy instrument that might correct this distortion. In this setup, there are two types of instruments that lend themselves for dealing with the input distortion. The first is an ad-valorem subsidy for material input use in variety production, such that the input price that variety producers face is $(1+s)\,\tilde{P}$. In the following we use $\theta := 1+s$. A subsidy implies $\theta < 1$, if $\theta > 1$ this means material input use is taxed. The second policy instrument is a subsidy of differentiated varieties used in assembly of the final good. This introduces a wedge between the price p set by producers of varieties and the "demand-price" $\tilde{p}$ faced by producers of the final good, such that $p = \tilde{p}/(1+t)$. In the following, we write $\tau := 1/(1+t)$, where a subsidy implies $\tau > 1$. If $\tau < 1$ this means the use of varieties is taxed. We emphasize that τ applies to the entire production of the composite good, including production for final consumption where there is no distortion. Hence, at first sight τ seems an unlikely candidate for a first best correction of the distortion. However, we shall see that in the present setup the two types policy instruments, θ and τ, are isomorphic.

Cost-minimizing production of a variety gives rise to minimum unit-cost $x := w^{\gamma}\left(\theta\tilde{P}\right)^{1-\gamma}$, and given demand as in equation (25), Bertrand pricing of variety producers yields

$$p = \mu x \quad \text{with} \quad \mu := \frac{\sigma}{\sigma-1} > 1. \tag{26}$$

Conditional input demands in production of varieties are

$$\ell = \gamma\frac{x}{w}q \quad \text{and} \quad m = (1-\gamma)\,\frac{x}{\theta\tilde{P}}q. \tag{27}$$

Free entry into the market for varieties implies zero profits, $(\mu-1)\,xq = fw$, which leads to the following equilibrium output per firm

$$q = \frac{1}{\mu-1}\frac{w}{x}f. \tag{28}$$

The variable labor use per firm then immediately follows as

$$\ell = \frac{\gamma}{\mu - 1} f. \tag{29}$$

Comparing this with equation (14), and noting that $\mu - 1 = \nu$, we recognize that the decentralized equilibrium features a first-best level of direct labor use per firm. Notice the different contexts in which the two terms $\mu - 1$ and ν are placed. When describing the first best in the previous section, we have used $\nu = 1/(\sigma - 1)$ as a measure of economies from the number of input varieties available for assembly of the aggregate good. Here, $\mu - 1 = 1/(\sigma - 1)$ additionally captures the degree of market power enjoyed by variety producers. The equilibrium number of firms follows from the full employment condition, $N(\ell + f) = L$, which implies

$$N = \left(\frac{\gamma}{\mu - 1} + 1 \right)^{-1} \frac{L}{f}. \tag{30}$$

Comparing with (15), we find that the decentralized equilibrium also features a first-best number of firms.

The equilibrium material input use, however, is distorted. It emerges as

$$m = \frac{1 - \gamma}{\gamma} \frac{w}{\theta \tilde{P}} \ell. \tag{31}$$

Inserting the markup pricing equation (26) into the aggregate price equation (24), we obtain

$$\tilde{P} = \theta^{\frac{1-\gamma}{\gamma}} N^{-\nu/\gamma} \left(\frac{\mu}{\tau} \right)^{1/\gamma} w, \tag{32}$$

which implies

$$\frac{m}{\ell} = \left(\frac{\tau}{\theta \mu} \right)^{1/\gamma} \frac{1 - \gamma}{\gamma} N^{\nu/\gamma}. \tag{33}$$

Applying the logic of roundaboutness in production developed above, we use (27) to write $A_m = Nm/Q$ as

$$A_m = (1 - \gamma) \left(\frac{w}{\theta \tilde{P}} \right)^{\gamma} \Big/ N^{\nu} = (1 - \gamma) \frac{\theta \mu}{\tau},$$

where the second equality uses (32) to replace $\left(\frac{w}{\theta\tilde{P}}\right)^{\gamma}$. The ratio of indirect to direct labor use in the variable input part of the technology is equal to

$$\frac{A_m}{1-A_m}=\frac{1-\gamma}{\gamma+\frac{\theta\mu}{\tau}-1}. \tag{34}$$

If $\theta\mu/\tau>1$, as in the laissez-faire case ($\theta=1$ and $\tau=1$) or with $\theta=\tau$, the decentralized equilibrium features a lower than socially optimal degree of roundaboutness in production. The total labor input per firm, inclusive of the fixed part, is equal to $\ell(1+\nu/\gamma)$, as in the social optimum. The ratio of indirect to total labor use (inclusive of the fixed labor input), taking on the value R^* as given in (19), may generally be written as

$$R=\left(\frac{1+\nu/\gamma}{A_m(1-A_m)^{-1}}+1\right)^{-1} \tag{35}$$

Obviously, if $A_m(1-A_m)^{-1}$ is lower than in the social optimum, then $R<R^*$ as well.

Finally, we may use equation (32) to derive the relative price of labor as

$$\frac{w}{\theta\tilde{P}}=\left(\frac{\tau}{\theta\mu}\right)^{\frac{1}{\gamma}}N^{\frac{v}{\gamma}}, \tag{36}$$

which we may compare to the shadow value of labor in the social planner's solution. From equations (15) and (18) we realize that $w\Big/(\tilde{P}\theta)=\left(\frac{\tau}{\theta\mu}\right)^{\frac{1}{\gamma}}\lambda^*$. If $\theta\mu/\tau>1$, as in the laissez-faire case ($\theta=1$ and $\tau=1$) or with $\theta=\tau$, then the relative price of labor lies below the shadow value of labor.

We may summarize our results on the decentralized equilibrium as follows:

Proposition 3 ***(a)*** *A decentralized laissez-faire equilibrium is characterized by a socially optimal level of employment in each firm as well as by an optimal number of firms in the market.* *(**b**) Compared to the social optimum, the material input use is lower than in the social optimum, and so is the degree of roundaboutness in production, causing an aggregate output loss as well as a consumption (welfare) loss.* ***(c)*** *In a subsidy/tax-ridden equilibrium the socially optimal level of consumption per capita is reached if the policy-wedges are such that* $\tau/\theta=\mu>1$.

Proof. **(a)** The first-best nature of ℓ as well as N is evidenced by equations (29) and (30) above. (**b**) Comparing (33) with equation (16), and setting $\tau = \theta = 1$ (laissez faire), we find that $m/m^* = \mu^{-1/\gamma} = [\sigma/(\sigma-1)]^{-1/\gamma} < 1$. As regards roundaboutness in production, see our remarks on (35) above. Similarly, inserting the above equations for ℓ, N and m (setting $\tau = \theta = 1$) into the equation for Q, and comparing with Q^* in (23) above, gives $Q/Q^* = \mu^{-(1-\gamma)/\gamma} = [\sigma/(\sigma-1)]^{(1-\gamma)/\gamma} < 1$. Finally, real consumption is $C = Q - M = (1 - M/Q)Q$, where $M = mN$. Since $N = N^*$, and given the ratios m/m^* and Q/Q^* from above, we have $M/Q = \left(\frac{\tau}{\theta\mu}\right)(M^*/Q^*)$. Remembering that $M^*/Q^* = (1-\gamma)$ and $C^* = \gamma Q^*$, we have

$$C = \frac{1}{\gamma}\left[1 - \frac{\tau}{\theta\mu}(1-\gamma)\right]\left(\frac{\tau}{\theta\mu}\right)^{(1-\gamma)/\gamma} C^*. \tag{37}$$

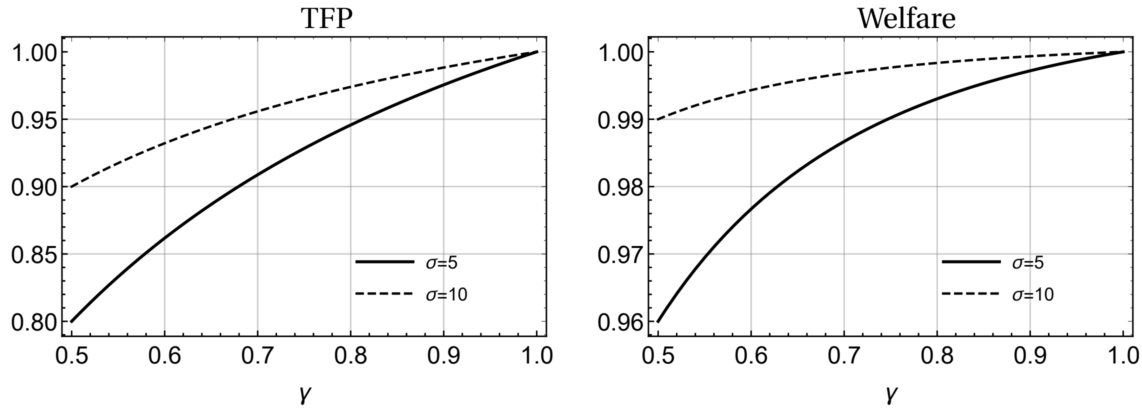
For $\tau = \theta = 1$, we have $0 < C/C^* < 1$. Note that for a given size of the labor force C also serves as a welfare measure. **(c)** For a subsidy/tax-ridden equilibrium, $\tau \neq 1$ and $\theta \neq 1$, it is immediately obvious from (37) that $\tau/\theta = \mu$ is a sufficient condition for $C = C^*$. A more thorough examination requires maximization of (37) with respect to τ/θ. It can be shown that the condition $\tau/\theta = \mu$ is also a necessary condition. ■

What is the magnitude of the welfare loss caused by the input distortion, absent any policy? The loss is measured by the ratio C/C^* in (37). For $\theta = \tau = 1$, we have $C/C^* = [(1 - 1/\mu + \gamma/\mu)/\gamma]\,\mu^{-(1-\gamma)/\gamma} < 0$. We are also interested in the loss in total factor productivity which is simply measured by $Q/Q^* = \mu^{-(1-\gamma)/\gamma}$. Figure 1 depicts these losses for values of $\gamma \in [0.5, 1]$, and for values of $\sigma \in \{5, 10\}$.[10] The losses are substantial, ranging, up to 4 percent for welfare and up to 20 percent for TFP.

Remember that $\tau > 1$ ($\tau < 1$) means subsidizing (taxing) production of differentiated goods, while $\theta < 1$ ($\theta > 1$) means subsidizing (taxing) the intermediate input use of the aggregate good. Hence, the optimal policy requires that the combined effect of the two types of policy amounts to net-subsidization at a rate equal to μ, which is the markup rate. This is as expected, since it is precisely this markup which is behind the input distortion. Moreover,

[10] In the appendix, we show that the share of intermediate inputs in gross output is between 40 and 60 percent. In an undistorted world, this share would reflect $1 - \gamma$.

Figure 1: Loss in total factor productivity and welfare induced by the input distortion



Legend: The curves depict the loss in total factor productivity (TFP) and consumption caused by the input distortion as a function of the labor cost share γ for different values of the elasticity of substitution.

it is unsurprising that subsidizing *intermediate input use* should be a first-best instrument to correct the input distortion. What is remarkable, however, is that the two types of policies considered here are equally suitable as a first-best correction of this distortion. Specifically, subsidizing *production* of differentiated goods, is an equally suitable first-best policy. After all, this policy effectively subsidizes production of the aggregate good irrespective of whether such production takes place for intermediate input use, or for final consumption, whereas the distortion is present only in intermediate input use.

Two things are important to understand this result. First, given perfect competition in aggregate goods production, subsidization of varieties used in assembly fully feeds into a lower price of this good. And secondly, while this lower price applies to both production and consumption (which is not distorted to start with), it does not distort consumption since by assumption all labor income is spent on this good. Moreover, we have implicitly assumed that the revenue needed to finance the subsidy is raised in a lump-sum (i.e., non-distortionary) way. One might wonder about the subsidy bill in the above analysis. The reason why this bill never showed up simply has to do with the fact that we did not approach consumption from the income side of the household sector. Instead, we have identified real consumption directly as what is left from aggregate output after taking account of intermediate input use.

What is the trade-off behind the optimal policy $\tau/\theta = \mu$? Consider a rise in τ/θ. This has two opposing effects on aggregate welfare C in (37). First, due to $M/Q = \left(\frac{\tau}{\theta\mu}\right)(M^*/Q^*)$ it increases the *share* of aggregate output Q devoted to intermediate input use, thus lowering the share available for consumption. But secondly, it also increases the *level* of aggregate output, due to $Q = [\tau/(\theta\mu)]^{(1-\gamma)/\gamma}\, Q^*$. Thus, we have two opposing forces emanating from the policy instrument τ/θ, and $\tau/\theta = \mu$ ensures an optimal trade-off between these two forces.

The industrial economists's immediate response to upstream markup pricing is vertical integration. It should be obvious that this is not a viable solution to our input distortion. There simply is no single material input supplier that the variety producer could possibly identify for vertical integration. Material inputs are bought on perfectly competitive markets. Moreover, the input-output linkage implies that it is the variety producer's own markup pricing which is at the heart of the problem.

3.3 Wage tax and wage setting

In the decentralized equilibrium, the material input intensity of variety production is governed by the price of the aggregate good relative to the wage rate; see equation (31). It is tempting to argue that a wage tax might be an equally suitable policy instrument to correct the input distortion. In a similar vein, one might argue that a wage setting environment leading to a wage that lies above the opportunity cost constitutes an offsetting distortion, mitigating the need for policy intervention or potentially even calling for a tax, rather than a subsidy on material input. It turns out that in our single-sector setting both arguments, while plausible in partial equilibrium, are wrong in general equilibrium.

Suppose that the government introduces an ad-valorem wage tax ϕ generating a wedge between w, the wage earned by workers, and $\tilde{w}$, the price for labor paid by variety producers:

$$\tilde{w} = (1+\phi)\, w. \tag{38}$$

With homogeneous labor, it seems reasonable to assume that that wage tax is uniformly imposed on variable and fixed labor input. A uniform wage tax does not distort the allocation between the variable and fixed type of labor use and, therefore, also does not distort the number of firms. The minimum unit cost in production of a variety is $x = \tilde{w}^{\gamma}\left(\theta\tilde{P}\right)^{1-\gamma}$. In *partial equilibrium*, i.e., for a *given* price of the aggregate good $\tilde{P}$, the wage tax clearly is an incentive to increase the material input intensity of production. Moreover, it is an incentive to increase output per firm according to $q = (\mu - 1)^{-1}\tilde{w}f/x$.

General equilibrium requires that we close the price-loop implied by the input-output linkage. A one percent increase in $1 + \phi$ in creases the minimum unit cost x as well as prices p and $\tilde{p}$ by γ percent. This follows from technology and markup pricing for varieties. Zero profits in assembly of the aggregate good then imply that the price of the aggregate good similarly increases by γ percent; see equation (24). Formally, the general equilibrium price adjustment of $\tilde{P}$ is governed by

$$\tilde{P} = \theta^{\frac{1-\gamma}{\gamma}} N^{-\nu/\gamma}\left(\frac{\mu}{\tau}\right)^{1/\gamma}\tilde{w}. \tag{39}$$

Comparing this with (32), we see that the relative price of the two inputs for variety production is invariant to the wage tax. So are the ratios $x/\tilde{w}$ and $x/\tilde{P}$ and, hence, input demands as well as outputs and the number of firms; see equations (27), (28) and (30). Put simply, in this model a wage tax doesn't entail any distortion.

We may summarize this result as follows:

Proposition 4 ***(a)*** *In our single-sector economy, a wage tax does not constitute a distortion and is, therefore, not suited to address the input distortion generated by monopolistic competition in the presence of an input-output linkage.* ***(b)*** *In an economy with monopolistic competition in the presence of an input-output linkage, an input distortion arises regardless of the underlying wage setting mechanism.*

Proof. Part (**a**) follows from the text above. Part (**b**) follows from considering alternative wage setting mechanisms (e.g., fair wages, efficiency wages, search and matching, or trade union wages) as deviations from the reference case of a perfect labor market. These deviations are isomorphic to the wage tax ϕ above. ∎

The intuition for this proposition is that with the input-output linkage and the price loop a wage tax doesn't affect any decision margin. We must, however, emphasize two caveats. The first is that we have assumed a completely inelastic labor supply. With elastic labor supply, a decision margin (e.g., consumption-leisure) arises which is affected by a wage tax as well as by wage setting mechanisms. The second is that our model has but one sector. In contrast to the input distortion, the entry distortion that arises in multi-sector settings interacts with wage setting mechanisms. One might also wonder about tax revenue, or the rents generated by wage setting. As with the subsidy instruments considered above, the reason why we need not consider tax revenue here is that in our approach we do not treat consumption from the income side of the household sector. Instead, we identify real consumption directly as what is left from aggregate output after taking account of intermediate input use.

3.4 Extensions

3.4.1 Fixed input in terms of the final good

In the above model, input-output linkages are restricted to the variable input part of production activities. From an empirical perspective, in many cases technology is such that intermediate inputs also loom large in the fixed input activities. We now develop a version of the model where the fixed input into production requires f quantities of the composite good rather than f units of labor. Except for this modification, the model remains as in the baseline case above.

Social optimum. The social planner maximizes $C = Q - M$ by choosing ℓ, m, and N, as above, but the resource constraint now reads as $L \leq N\ell$, whereas the total material input (in terms of the composite good) in production of varieties is $M = N(m + f)$. The corresponding Lagrangian reads as

$$\pounds = Q - M - \lambda [N\ell - L]. \tag{40}$$

As in the baseline, the conditions on ℓ^* and m^* emerge as

$$\ell^* = \gamma \frac{Q^*}{\lambda^* N^*} \text{ and } m^* = (1-\gamma)\frac{Q^*}{N^*}. \tag{41}$$

The condition on N^* reads as

$$(\nu + 1) Q^* - M^* = \lambda^* L, \tag{42}$$

where M^* now contains total use of the composite good in variable and fixed input into production. The resource constraint requires $\ell^* = L/N^*$.

Employing the production function in the condition on m^*, using relative input demand, and observing the resource constraint, we may see that the Lagrange parameter emerges as in the baseline case above:

$$\lambda^* = (N^*)^{\frac{\nu}{\gamma}}. \tag{43}$$

Using the condition on m^* in the condition on N^*, we obtain

$$(\nu + \gamma) Q^* = f N^* + \lambda^* L. \tag{44}$$

Using the condition on ℓ^* and using the resource constraint to get rid of Q^* and N^* in the above expression, we obtain a second relationship between λ^* and N^*:

$$\lambda^* = \frac{\gamma}{\nu}\frac{f}{L} N^*. \tag{45}$$

Combining the two expressions, we can solve for N^* as

$$N^* = \left(\frac{\nu}{\gamma}\frac{L}{f}\right)^{\frac{\gamma}{\gamma-\nu}}. \tag{46}$$

Optimal material input and labor input emerge as

$$m^* = \frac{1-\gamma}{\gamma}\lambda^* \ell^* = \frac{1-\gamma}{\nu} f \quad \text{and} \quad \ell^* = (\lambda^*)^{-1}\frac{\gamma f}{\nu}. \tag{47}$$

Finally, we check the parameter restriction implied by viability, $Q^* \geq N^*(m^* + f)$. Using the condition on m^* to substitute out Q^*, we obtain

$$\frac{N^* m^*}{(1-\gamma)} \geq N^* m^* \left(1 + \frac{f}{m^*}\right). \tag{48}$$

Employing the solution for m^*, we may rewrite this as $\gamma \geq \nu$.

Decentralized equilibrium with policy intervention. While conditional demands of the final good producer and the price of the composite good are the same as in the baseline case, the zero profit condition now implies

$$q = \frac{\theta \tilde{P}}{x} \frac{f}{\mu - 1}. \tag{49}$$

Using this expression to substitute out firm size from conditional demands for material input and labor input, respectively, we obtain

$$m = \frac{1-\gamma}{\mu - 1} f \quad \text{and} \quad \ell = \left(\frac{w}{\theta \tilde{P}}\right)^{-1} \frac{\gamma f}{\mu - 1}, \tag{50}$$

where

$$\frac{w}{\theta \tilde{P}} = \left(\frac{\tau}{\mu \theta}\right)^{\frac{1}{\gamma}} N^{\frac{\nu}{\gamma}}. \tag{51}$$

Allocational efficiency and optimal policy. In this version of the model, material input is efficient, while in the absence of policy intervention ($\tau = \theta = 1$), labor input is inefficiently large and – by implication – the number of firms is inefficiently low. As in the baseline, the optimal policy to offset this input distortion is

$$\frac{\tau}{\theta} = \mu > 1. \tag{52}$$

We now explore this implications for roundaboutness of production. The direct labor input coefficient reads as $A_\ell = \ell N / Q$. The direct material input coefficients for production fixed input are, respectively,

$$A_{mp} = \frac{mN}{Q} \text{ and } A_{mf} = \frac{fN}{Q}. \tag{53}$$

The combined material input, $A_{mp} + A_{mf}$, requires further material input use in the amount of $A_{mp}(A_{mp} + A_{mf})$. Reiterating, we obtain a total material input use per unit of final output equal to $(A_{mp} + A_{mf})(1 - A_{mp})^{-1}$. Note that $A_{mp} + A_{mf}$ requires no further (indirect) material input use on account of the fixed cost. The indirect labor use (embodies in material inputs) is $A_\ell(A_{mp} + A_{mf})(1 - A_{mp})^{-1}$. For easier writing, we now measure roundaboutness by the indirect labor use relative to the direct labor use. This measure is

$$R_1 = (A_{mp} + A_{mf})(1 - A_{mp})^{-1}. \tag{54}$$

In the social optimum, we have $A^*_{mp} = 1 - \gamma$ and $A^*_{mf} = \frac{f}{m^*} A^*_{mp} = \nu$, which results in

$$R^*_1 = \frac{1 - \gamma + \nu}{\gamma}. \tag{55}$$

In the decentralized equilibrium without policy intervention, we have $A_{mp} = A^*_{mp}/\mu$ and thus $A^*_{mf} = A^*_{mf}/\mu$. Summing up, we obtain

$$R_1 = \frac{1 - \gamma + \nu}{(\mu - 1) + \gamma} < R^*_1 \tag{56}$$

Hence, in the decentralized equilibrium the degree of roundaboutness is too small, as in the baseline case.

We can also quantify the implications for total factor productivity and consumption. With respect to TFP $= Q/L$, we have

$$\text{TFP} = \left(\frac{N}{N^*}\right)^{\nu+1} \left(\frac{\ell}{\ell^*}\right)^{\gamma} \text{TFP}^*, \tag{57}$$

where $(N/N^*)^{\nu+1} < 1$ and $(\ell/\ell^*)^\gamma > 1$ reflect the effects of, respectively, an inefficiently small number of firms and an inefficiently large labor input into production. On net, we have

$$\text{TFP} = \mu^{-\frac{1-\gamma+\nu}{\gamma-\nu}}\text{TFP}^*. \tag{58}$$

Compared to the baseline above, the TFP discrepancy is larger.[11]

Turning to real (per capita) consumption, we have

$$C = \mu^{-\frac{1-\gamma+\nu}{\gamma-\nu}} \frac{1 - \frac{1-\gamma}{\mu}}{\gamma} C^*. \tag{59}$$

The distortion in the *share* of the composite output used as material input is the same as in the baseline. Hence, the larger discrepancy in TFP compared to the baseline directly translated into a larger discrepancy in real consumption.

We may summarize the result of this extension in the following proposition:

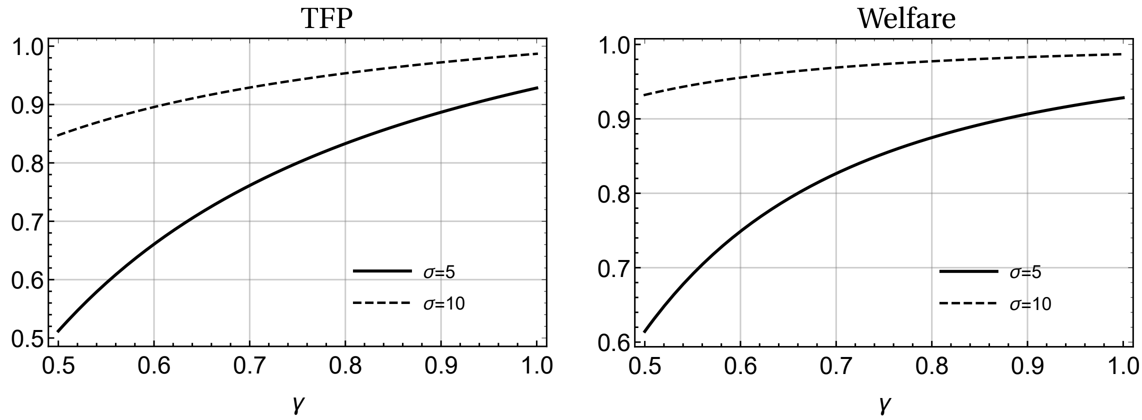
Proposition 5 *If the input-output linkage affects the fixed cost on the same footing as it does the variable inputs, then the decentralized market equilibrium has the following properties, compared to the social optimum: (**a**) The equilibrium level of material inputs is undistorted, relative to the social optimum, but the labor input is inefficiently large. As in the baseline case, the degree of roundaboutness is suboptimally low. (**b**) The equilibrium number of firms is no longer undistorted, but is suboptimally low. (**c**) The welfare loss caused by the input distortion is now larger than in the baseline case; so is the loss in terms of total factor productivity.*

Proof. All parts follow from the text above. ■

Part (**a**) of the proposition is intuitive. The distortion works in the same direction regarding the input intensity as in the baseline case, but this time due to an inefficiently large amount of labor, rather than too little material input use. Part (**b**) implies a new channel through which the input distortion causes inefficiency, viz. the number of firms. Intuitively, this leads to a larger welfare loss, relative to the benchmark case. Figure 2 shows that the loss

[11] This follows from noting that $\frac{1-\gamma+\nu}{\gamma-\nu} > \frac{1-\gamma}{\gamma}$.

Figure 2: Loss in TFP and welfare with a fixed input in terms of the composite good



Legend: The curves depict the loss in total factor productivity (TFP) and consumption caused by the input distortion as a function of the labor cost share γ for different values of the elasticity of substitution when the fixed input requires the final good.

is now very large indeed, with a maximum value of almost 40 percent in the case where $\sigma = 5$. Note that a lower σ implies a high variety effect, meaning that the loss in the number of firms is now felt more strongly in welfare terms.

3.4.2 Heterogeneous firms

We now allow differentiated good producers to differ in terms of their productivity levels φ such that

$$q(\varphi) = \varphi \left(\frac{l(\varphi)}{\gamma} \right)^{\gamma} \left(\frac{m(\varphi)}{1-\gamma} \right)^{1-\gamma}. \tag{60}$$

Let productivities be distributed according to the cumulative density function $G(\varphi)$. With heterogeneous firms, real consumption is given by

$$C = Q - M = \left[N \int_{\varphi_c}^{\infty} q\left(\varphi\right)^{\frac{\sigma-1}{\sigma}} \mathrm{d}G(\varphi) \right]^{\frac{\sigma}{\sigma-1}} - N \int_{\varphi_c}^{\infty} m\left(\varphi\right) g\left(\varphi\right) \mathrm{d}G(\varphi) \tag{61}$$

where N now denotes the mass of entrants and φ_c denotes the cutoff productivity level. The following observations stand out from equation (61):

1. For a given mass of entrants N and a given productivity cutoff φ_c, the first term is in-

creasing in $l(\varphi)$. It is disciplined, however, by the resource constraint of the primary factor (labor), i.e., (*i*) for a given cutoff φ_c, a higher $l(\varphi)$ implies a lower N, and (*ii*) for a given N, a higher $l(\varphi)$ implies a higher cutoff φ_c.

2. Material input $m(\varphi)$ can be changed *independently* of labor input $l(\varphi)$. The corresponding "resource constraint" is $M \leq Q$, but $m(\varphi)$ is not directly bound by the resource constraint of the primary factor (labor).

3. For given N and φ_c, the first term rises in $m(\varphi)$ (higher output), while the second falls in $m(\varphi)$ (higher material input).

These observations imply that the planner's real consumption maximization problem can be decomposed into two stages. In the first stage, aggregate output Q is maximized for given $\{m(\varphi)\}$. In the second stage, the planner finds $\{m(\varphi)\}$ that maximizes real consumption C for given N, φ_c, and $\{l(\varphi)\}$. The market, however, maximizes aggregate revenue for given $\{m(\varphi)\}$ in the first stage. In the second stage, material input is chosen to minimize firms' cost, which is guided by the relative price of labor

$$\frac{w}{\theta\tilde{P}} = \left(\frac{\tau}{\theta\mu}\right)^{\frac{1}{\gamma}} N^{\frac{\nu}{\gamma}} \left(\int_{\varphi_c} \varphi^{\sigma-1}\mathrm{d}G(\varphi)\right)^{\frac{\nu}{\gamma}}. \tag{62}$$

Consider the *first stage*. Dhingra and Morrow (2019) show that *for given* $\{m(\varphi)\}$, the market and optimal allocations N, φ_c, and $\{l(\varphi)\}$ can be expressed as solutions to

$$\max R = N \int_{\varphi_c} u'(q(\varphi))\, q(\varphi)\, dG(\varphi) \text{ s.t. resource constraint (market) and}$$
$$\max Q = N \int_{\varphi_c} u(q(\varphi))\, dG(\varphi) \text{ s.t. resource constraint (optimum).}$$

With CES, we have $u(q(\varphi)) = q(\varphi)^{\frac{\sigma-1}{\sigma}}$ and $u'(q(\varphi))\, q = \rho q(\varphi)^{\frac{\sigma-1}{\sigma}}$. Revenue maximization is perfectly aligned with welfare maximization such that the mass of entrants N, the entry cutoff φ_c, and variable labor input per firm $\{l(\varphi)\}$ are efficient.

Turn now to the *second stage*. As the mass of entrants N and the cutoff φ_c are efficient, markup pricing is the only source of distortion of the price of the aggregate good. Hence, firms use too little material input, as in the case of homogeneous firms. Again, a subsidy on the purchase of intermediate inputs that exactly offsets the mark-up yields the efficient allocation of material input.

4 Country borders and trade costs

We now explore how the two countries cooperatively set their policies if they are separated by borders. The presence of borders gives rise to two additional margins. First, the outputs Y_h and Y_f have to be distributed among the two countries for production of the nontrable final goods Q_h an Q_f. We write Y_d and Y_m for conditional domestic and import demand, respectively. Second, the government may now subsidize the use of the domestically produced and the imported country-specific aggregates differently by τ_d and τ_m. Due to symmetry, we suppress the country indices, albeit we have to be bare in mind that now L and N refer to endowments and the number of firms at the country level, while in section 3 these variables refer to the integrated world economy.

4.1 Full set of policies

We suppose that the governments have all policy instruments τ_d, τ_m, and θ at their disposal.

Social optimum. The socially optimal allocation follows from solving

$$\max_{\ell,m,N,Y_m} 2\left[((Y-\delta Y_m)^\rho + Y_m^\rho)^{\frac{1}{\rho}} - Nm\right] \tag{63}$$

subject to the production functions and the resource constraints. The 2 appears in this expression as we deal with two symmetric countries. Compared to the baseline, there is a new margin: the optimal choice of the imported quantity Y_m. In the statement of the maximization problem, we already made use of the goods market clearing condition. The corresponding Lagrangian is

$$\pounds = \left((Y - \delta Y_m)^\rho + Y_m^\rho \right)^{\frac{1}{\rho}} - Nm - \lambda \left[N \left(l + f \right) - L \right]. \tag{64}$$

The first-order conditions for ℓ and m emerge as

$$\ell^* = \gamma \frac{\Gamma^* Y^*}{\lambda^* N^*} \text{ and } m^* = (1 - \gamma) \frac{\Gamma^* Y^*}{N^*}, \tag{65}$$

where

$$\Gamma^* \equiv \left(\frac{Q^*}{Y_d^*} \right)^{1-\rho}. \tag{66}$$

The first-order condition for imports Y_m is

$$-\rho \left(Y^* - \delta Y_m^* \right)^{\rho-1} \delta + \rho \left(Y_m^* \right)^{\rho-1} = 0 \Rightarrow Y_m^* = \frac{\delta^{\frac{1}{\rho-1}}}{1 + \delta^{1+\frac{1}{\rho-1}}} Y^*. \tag{67}$$

Evoking goods market clearing, relative import demand emerges as

$$\frac{Y_m^*}{Y_d^*} = \delta^{-\sigma}. \tag{68}$$

Intuitively, trade costs drive a wegde into import and domestic demand. Using equation (68) and goods market clearing, we obtain

$$\Gamma^* Y^* = \left(\frac{Q^*}{Y_d^*} \right)^{1-\rho} \left(Y_d^* + \delta Y_m^*) \right) = Q^*. \tag{69}$$

Hence, the conditions on ℓ and m in equation (65) collapse to their counterparts in the baseline. The first-order condition on N is

$$\Gamma \left(\nu + 1 \right) Y^* - N^* m^* = \lambda^* L. \tag{70}$$

Using the conditions on m and ℓ to substitute out m^* and λ^*, respectively, from equation (70) and evoking the resource constraint, we can solve for variable labor input as

$$\ell^* = \frac{\gamma}{\nu} f. \tag{71}$$

Variable labor input into production of differentiated varieties is not affected by the presence of borders. Employing the resource constraint, the optimal number of firms emerges as

$$N^* = \frac{1}{\frac{\gamma}{\nu}+1} \frac{L}{f}. \tag{72}$$

This expression is structurally equivalent to the one obtained in the integrated world equilibrium, the difference being that here L refers to a single country's labor endowments, while in the integrated world equilibrium L represents world labor endowments.

Using again the first-order condition on m to substitute out m^* from equation (70) and employing equation (71) and the resouce constraint, the Lagrange parameter now reads as

$$\lambda^* = \Gamma^{\frac{1}{\gamma}} \left(N^*\right)^{\frac{\nu}{\gamma}}, \tag{73}$$

where, in constrast to the baseline, N^* refers to the number of firms in one of the countries. The term $\Gamma^{\frac{1}{\gamma}}$ appears as a relaxation of the resource constraint not only affects consumption through the domestic, but also through the imported country-specific aggregate. Employing equation (68), we obtain

$$\Gamma = \left(\frac{Q^*}{Y_d^*}\right)^{1-\rho} = \left(\frac{\left((Y_d^*)^\rho + (Y_m^*)^\rho\right)^{\frac{1}{\rho}}}{Y_d^*}\right)^{1-\rho} = \left(1 + \left(\frac{Y_m^*}{Y_d^*}\right)^\rho\right)^{\frac{1-\rho}{\rho}} = \left(1+\delta^{1-\sigma}\right)^\nu. \tag{74}$$

In the absence of trade costs ($\delta = 1$), Γ collapses to 2^ν, and the Lagrange parameter reads as $\lambda^* = (2N^*)^{\frac{\nu}{\gamma}}$. In general, the Lagrange multiplier is decreasing in trade costs.

Rearranging terms in equation (74), and employing equation (68), optimal aggregate output emerges as

$$Q^* = \left(1+\delta^{1-\sigma}\right)^{\frac{1}{\rho}} Y_d^* = \left(1+\delta^{1-\sigma}\right)^\nu Y^*, \tag{75}$$

where $Y^* = (N^*)^{\nu+1} q^*$. In the absence of trade costs ($\delta = 1$), aggregate output is $Q^* = 2^\nu Y^* < 2Y^*$. The inequality is a consequence of external economies of scale. With scale economies, restricing labor mobility results results in lower aggregate output, even in the absence trade costs. Employing the production functions and using the conditions on ℓ and m, we obtain

$$Y^* = \frac{f}{\nu}\Gamma^{\frac{1-\gamma}{\gamma}} (N^*)^{\frac{\nu}{\gamma}+1} \quad \text{and} \quad Q^* = \frac{f}{\nu}\Gamma^{\frac{1}{\gamma}} (N^*)^{\frac{\nu}{\gamma}+1}. \tag{76}$$

It immediately follows from the condition on m and equation (69) that the share output used as input into production of differentiated varieties reads as $M^*/Q^* = 1-\gamma$, as in the baseline. Hence, consumption is $C^* = \gamma Q^*$.

Decentralized equilibrium with policy intervention. In decentralized equilibrium, conditional demand for inputs at the level of the differentiated good producers are the same as the basic setting; see equation (27). Moreover, their pricing behavior and the zero profit condition are the same, giving rise to the following solution in decentralized equilibrium; see equations (29) to (31):

$$\ell = \frac{\gamma}{\mu-1}f, N = \left(\frac{\gamma}{\mu-1}+1\right)^{-1}\frac{L}{f}, \text{ and } m = \frac{1-\gamma}{\gamma}\frac{w}{\theta\tilde{P}}\ell. \tag{77}$$

The new margin is that the final good producer has to combine the two country-specific aggregates. Cost minimizing behavior of the final good producer implies that relative conditional input demand for imports is given by

$$\frac{Y_m}{Y_d} = \left(\frac{\tau_d}{\tau_m}\delta\right)^{-\sigma}, \tag{78}$$

where on top of trade costs, the relative policies drive a wedge into these two quantities. The price of the final good is given by

$$\tilde{P} = \left(N\tilde{p}_d^{1-\sigma} + N\tilde{p}_m^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = N^{-\nu}\frac{\mu w^\gamma \left(\theta\tilde{P}\right)^{1-\gamma}}{\tau_d}\left(1+\left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}. \tag{79}$$

Solving the loop in equation (79) for $\tilde{P}$, the relative price of labor emerges as

$$\frac{w}{\theta \tilde{P}} = \left(\frac{\tau_d}{\theta\mu}\right)^{\frac{1}{\gamma}} \left[\left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right) N\right]^{\frac{v}{\gamma}}. \tag{80}$$

The first term is exactly the same as in the integrated world equilibrium. The second term highlights the effects of the trade costs and relative policies. In the absence of trade costs ($\delta = 1$) and policy differentials ($\tau_d = \tau_m$), it collapses to $(2N)^{\frac{\nu}{\gamma}}$. In the absence of policy differentials, the relative price of labor is decreasing in trade costs.

Employing the production function and evoking goods market clearing, aggregate output emerges as

$$Q = \left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)^{\frac{1}{\rho}} Y_d = \frac{\left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)^{\frac{1}{\rho}}}{1 + \delta^{1-\sigma}\left(\frac{\tau_d}{\tau_m}\right)^{-\sigma}} Y, \tag{81}$$

where we may rewrite

$$\left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)^{\frac{1}{\rho}} = \left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)\left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)^{\nu}$$

in order to faciliate easy comparison to Q^*. In the absence of policy differentials ($\tau_d = \tau_m$), the expression collapses to $Q = \left(1 + \delta^{1-\sigma}\right)^{\nu} Y$, which resembles equation (75). Employing the production functions and using conditional input demands for ℓ and m, we obtain

$$Q = \frac{f}{\nu}\left(\frac{\tau_d}{\theta\mu}\right)^{\frac{1-\gamma}{\gamma}} \frac{1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}}{1 + \delta^{1-\sigma}\left(\frac{\tau_d}{\tau_m}\right)^{-\sigma}} \left(1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}\right)^{\frac{v}{\gamma}} N^{\frac{\nu}{\gamma}+1}. \tag{82}$$

In the absence of policy differentials, aggregate output reads as $Q = \left(\frac{\tau_d}{\theta\mu}\right)^{\frac{1-\gamma}{\gamma}} Q^*$, as in the baseline. Again employing the production functions and using conditional input demands for ℓ and m, the share of output used as input into production emerges as

$$\frac{M}{Q} = (1-\gamma)\frac{\tau_d}{\theta\mu}\frac{1 + \delta^{1-\sigma}\left(\frac{\tau_d}{\tau_m}\right)^{-\sigma}}{1 + \left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}}. \tag{83}$$

In the absence of policy differentials, the share collapses to $(1-\gamma)\,\frac{\tau_d}{\theta\mu}$, again as in the baseline. Consumption is given by $C = Q(1 - M/Q)$.

Comparing the decentralized equilibrium to social optimum, two observations stand out. First, the allocation of the country-specific aggregate to markets Y_m/Y_d is optimal if the government does not condition the subsidy on the country of origin: $\tau_d = \tau_m$. Second, conditional on $\tau_d = \tau_m$, the decentralized equilibrium replicates social optimum if the government subsidizes either the use of the country-specific aggregates in the production of the final good or the use of the final good in the production of differentiated varieties to offset the markup. The following proposition straightforwardly generalizes Proposition 3 to the case of borders.

Proposition 6 ***(a)*** *In a cooperative setting with borders and symmetric countries, a decentralized laissez-faire equilibrium is characterized by a socially optimal level of employment in each firm as well as by an optimal number of firms in each market.* ***(b)*** *Compared to the social optimum, the material input use is lower than in the social optimum, causing an aggregate output loss as well as a consumption (welfare) loss.* ***(c)*** *In a subsidy/tax-ridden equilibrium the socially optimal level of consumption per capita is reached if the policy-wedges are such that* $\frac{\tau_d}{\theta} = \mu$ *and* $\tau_d = \tau_m$.

Proof. **(a)** The first-best nature of ℓ and N follow from comparing (77) to (71) and (72). **(b)** Follows from noting that $\frac{m}{m^*} = \left(\frac{\tau_d}{\theta\mu}\right)^{\frac{1}{\gamma}} \left(\frac{1+\left(\frac{\tau_d}{\tau_m}\delta\right)^{1-\sigma}}{1+\delta^{1-\sigma}}\right)^{\nu/\gamma}$. **(c)** A sufficient condition for $C = C^*$ is $\frac{\tau_d}{\theta} = \mu$ and $\tau_d = \tau_m$. ■

An important corollary to part (c) of the proposition is that trade costs do not affect optimal policies.

4.2 Restriction to trade policy

We continue to assume that the two symmetric countries set their policies cooperatively, but now we assume that they are restricted to the use of trade policy measures, while domestic subsidies are not available.[12] With cooperation, the terms-of-trade externality is internalized. While in the standard two-country, single-sector CES setting, free trade is optimal, in our setting with an input-output linkages, cooperative trade policy might be used as a second-best instrument to address the input distortion. The intuition is the following. An import subsidy lowers the price of the imported country-specific aggregate, and this reduction is partly passed on the price of composite good, which, in turn, alleviates the distortion of the relative price of labor. The welfare cost of the import subsidy, however, is a distortion of the relative price of the imported country-specific aggregate. In the presence of an import subsidy, the final good producer uses too much of the imported country-specific aggregate relative to the domestic one.

To formalize the argument, we consider a setting in which governments take the behavior of all types of firms and $\tau_d = \theta = 1$ as given and cooperatively choose τ_m to maximize consumption $C = Q - M$. For the sake of the clarity of the argument, we abstract from trade costs ($\delta = 1$). Equations (76) and (82) imply that in this restricted setting, aggregate output emerges as

$$Q = \mu^{-\frac{1-\gamma}{\gamma}} \frac{\left(1+\tau_m^{\sigma-1}\right)^{\frac{v}{\gamma}+1}}{1+\tau_m^{\sigma}} Q^*. \tag{84}$$

The effect of import subsidy on aggregate output seems to be ambiguous. Totally differentiating the above expression with respect to Q and τ_m, we obtain

$$\frac{d\ln Q}{d\ln \tau_m} = \tau_m^{\sigma-1}\left[\frac{\frac{1}{\gamma}+\sigma-1}{1+\tau_m^{\sigma-1}} - \frac{\sigma\tau_m}{1+\tau_m^{\sigma}}\right]. \tag{85}$$

Evaluated at $\tau_m = 1$, we have

$$\left.\frac{d\ln Q}{d\ln \tau_m}\right|_{\tau_m=1} = \frac{1-\gamma}{2\gamma} > 0. \tag{86}$$

Hence, a *small* import subsidy ($\tau_m > 1$) raises aggregate output.

Employing the production functions and using the expressions for labor and material in-

[12] The Agreement on Subsidies and Countervailing Measures disciplines the use of production subsidies.

put, the share of output used as material input emerges as

$$\frac{M}{Q}=\frac{1-\gamma}{\mu}\frac{1+\tau_m^{\sigma}}{1+\tau_m^{\sigma-1}}. \tag{87}$$

It is easy to check that the share of output used as input into production is increasing in an import subsidy. By mirror image, an import subsidy (larger τ_m) shrinks the share of output used for consumption.

In order to demonstrate that the net effect a *small* import subsidy on real consumption is positive, we totally differentiate the share of output used for final consumption:

$$\frac{d\ln\left(1-\frac{M}{Q}\right)}{d\ln\tau_m}=-\frac{1-\gamma}{\mu}\frac{\sigma\left(\tau_m-1\right)+1+\tau_m^{\sigma}}{\left(1+\tau_m^{\sigma-1}\right)^2}\frac{\tau_m^{\sigma-1}}{1-\frac{1-\gamma}{\mu}\frac{1+\tau_m^{\sigma}}{1+\tau_m^{\sigma-1}}}<0. \tag{88}$$

Evaluating at $\tau_m=1$ and using equation (86), we obtain

$$\left.\frac{d\ln C}{d\ln\tau_m}\right|_{\tau_m=1}=\frac{1-\gamma}{\gamma}\frac{\mu-1}{2\left(\mu-1+\gamma\right)}>0. \tag{89}$$

Thus, a *small* cooperative import subsidy is consumption (welfare) enhancing. Intuitively, the effect of a small import subsidy on consumption becomes negligible when either $\gamma\to 1$ or $\mu\to 1$ ($\rho\to 1$).

We may summarize our result as follows:

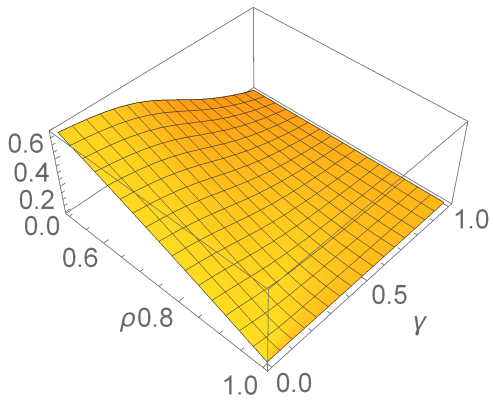
Proposition 7 *A cooperative import subsidy is the second-best policy to address the input distortion generated by monopolistic competition in the presence of an input-output linkage.*

Proof. See equation (89). ■

Recall that in the standard setting without the input-output linkage, free trade is optimal. Our result has important implications for non-cooperative optimal trade policy. The input distortion runs counter to the standard terms-of-trade externality. Hence, in a setting with an input-output linkage, the optimal tariff is lower than in the standard setting.[13]

[13] Our numerical simulations suggest that the input distortion has the potential to dominate the standard

Figure 3: Optimal cooperative import subsidy in the absence of domestic policies



The graph shows the optimal cooperative import subsidy $|t| = 1 - 1/\tau_m$ in the absence of domestic policies ($\tau_d = \theta = 1$). $\rho = (\sigma - 1)/\sigma \in (0.5, 1)$ is a transformation of the elasticity of substitution. $\gamma \in (0, 1)$ is labor cost share in production.

Figure 3 illustrates the consumption maximizing cooperative import subsidies $|t| = 1 - 1/\tau_m$ in (ρ, γ)-space.[14] The smaller the labor cost share γ and/or the elasticity of substitution (or: ρ), the larger the optimal cooperative import subsidy. For $\sigma = 5$ ($\rho = 0.8$) and $\gamma = 0.5$, the optimal cooperative import subsidy amounts to 7.4 percent.

We can also quantify the welfare consequences of prohibiting the use of domestic policies. In the presence of an import tariff, the consumption (welfare) discrepancy emerges as

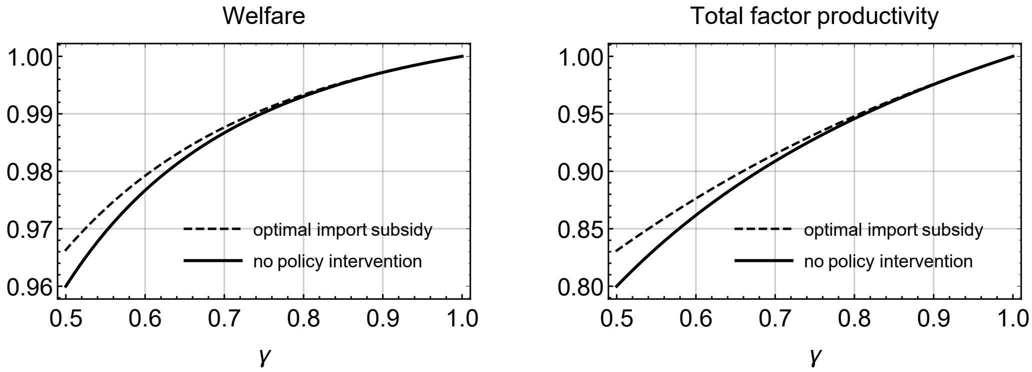
$$\frac{C}{C^*} = \mu^{-\frac{1-\gamma}{\gamma}} \frac{\left(1 + \tau_m^{\sigma-1}\right)^{\frac{v}{\gamma}+1}}{1 + \tau_m^{\sigma}} \frac{1 - \frac{1-\gamma}{\mu} \frac{1+\tau_m^{\sigma}}{1+\tau_m^{\sigma-1}}}{\gamma}. \tag{90}$$

Figure 4 illustrates the TFP and and consumption discrepancies as a function of γ for $\sigma = 5$, evaluated at the optimal cooperative import subsidy. Compared to the situation without any policy intervention, the use of an optimal cooperative import subsidy has only small effects on the discrepancies. For $\gamma = 0.5$, optimally subsidizing imports shrinks the consump-

terms-of-trade considerations under certain parameter constellations, turning the optimal non-cooperative trade policy in an import subsidy, even in the absence of an entry distortion inherent to multi-sector models; see Caliendo et al. (2017).

[14]The optimal cooperative import subsidy is determined by the first-order condition of the welfare maximization problem: $\frac{\partial C}{\partial \tau_m} = 0$. In order to compute the policy-induced change in consumption, we employ equations (85) and (88).

Figure 4: Discrepancies with and without optimal cooperative import subsidies



Legend: The curves depict the loss in total factor productivity (TFP) and welfare caused by the input distortion as a function of the labor cost share γ for different values of the elasticity of substitution. The *solid line* refers to a situation without policy intervention, the *dashed line* to a situation in which the optimal cooperative import subsidy is employed.

tion and TFP loss induced by the input distortion, respectively, from 4 to 3.4 percent and from 20 to 17 percent.

5 Summary and conclusions

Modern trade literature emphasizes that product differentiation is an important source of consumer welfare, while firms gain from the availability of differentiated intermediate inputs. But product differentiation comes at the cost of market power and prices above marginal cost, which in turn is a source of welfare loss. There is a voluminous literature addressing the various distortions that the twin feature of market power and product differentiation may entail. In general, the distortions arising from monopolistic competition are well understood. Most of the literature focuses on environments of monopolistic competition. Oftentimes, it also assumes an input-output linkage, meaning that production of differentiated goods uses material inputs alongside primary inputs. In this paper we demonstrate that the combination of monopolistic competition and input-output linkages gives rise to a distortion that has so far gone unnoticed. The reason simple: markup pricing for goods means that prices of material inputs lie above their opportunity cost, which potentially distorts the mix of material and primary inputs – we speak of an input distortion. This causes a welfare loss over and above the potential loss deriving from distortions highlighted by existing literature.

We develop a stylized model that zooms in on this input distortion by assuming away all other potential distortions deriving from monopolistic competition. In particular, we assume a single sector and we model monopolistic competition based on the CES-version of love of variety. Following existing literature, we model the input output linkage by means of a Cobb-Douglas production function for differentiated varieties, using material inputs as well as labor. This nests the simpler case without input-output linkage and allows us to explore what the presence and strength of input-output linkages means for welfare and total factor productivity. We first provide a full description of the social optimum for a fully integrated world economy featuring imperfect competition in the presence of such an input-output linkage. In doing so, among other things we also analyze the optimal degree of roundaboutness in production. By roundaboutness, we mean that labor is used to facilitate goods consumption both, in a direct way and an indirect way through using material input in production. The social optimum of the integrated world economy presents a benchmark for our analysis of the production inefficiency caused by the input-output linkage in a decentralized market equilibrium. It also establishes a reference case against which we discuss the implications of trade in intermediate goods.

It is well known from the literature that the comparative statics of a monopolistic competition equilibrium much depends on whether or not the fixed cost relies on the same input bundle as the variable cost. In our baseline case we assume that the fixed cost arises in the form of labor (the primary input), whereas the variable cost arises from both labor and material inputs. In an extension, we allow for the input-output linkage to be present also in the fixed cost. In the baseline case, markup pricing leaves the labor input as well as the number of firms undistorted, relative to the social optimum, whereas the material input is used in a less than optimal level which is also responsible for a suboptimally low degree of roundaboutness. Importantly these deviations from the social optimum generate a sizable welfare loss. For plausible values of the key parameters, the welfare loss is in the vicinity of two to four percent, relative to the first best. The loss in total factor productivity is even larger, between

10 and 20 percent. In the alternative case where the input-output linkage extends to the fixed cost, we again find a distorted input mix, but this time it comes from a higher than optimal labor use, relative to material inputs which are now first-best. Interestingly, in this case we also find that there is a further channel through which the input distortion works out in the decentralized equilibrium, which is a lower than optimal number of firms. This, in turn, is responsible for a magnified welfare loss, which for plausible parameter values lies between 10 and 20 percent. In a further extension, we allow for Melitz-type heterogeneity among firms. It turns out that this does not alter our main results in any way.

What are suitable policies to address the input distortion? Intuitively, the first-best policy is to subsidize material input use in production of differentiated varieties. In principle, subsidizing production of differentiated varieties would also seem a suitable instrument to address the distortion, but one expects this to be a second-best policy since it does not directly target the distortion which lies with the input mix. We describe the decentralized market equilibrium simultaneously allowing for both of these policy instruments, and it turns out that they are perfect substitutes for each other. The reason is that in the integrated world equilibrium the output subsidy does not, in and of itself, involve any distortion. One would expect that a wage tax is similarly able to address the input distortion. However, we demonstrate that this is partial equilibrium intuition and that in general equilibrium the input-output linkage implies the wage tax is fully passed on to the price of material input. Therefore, it is unable to influence the input mix chosen by producers. Indeed, it turns out that in our stylized model a wage tax simply doesn't constitute a distortion and cannot, therefore, serve in offsetting the input distortion. The same applies for deviations from the benchmark model in the form of wage setting environments, instead of a perfectly competitive labor market. An important caveat here is that these conclusions would be altered in a model allowing for many sectors and/or endogenous labor supply.

The most interesting implications of the input distortion arise in a trading environment where countries have completely segmented labor markets and where trade is subject to trade barriers. We look at the simple case of two symmetric countries that may trade in intermediate inputs. We readdress the question of optimal cooperative policies in two different settings. The first is a setting where domestic subsidies and trade policies are available. In this setting, it is optimal to subsidize domestic and imported varieties in the same way in order avoid a distortion of relative import demand. The second is one where only trade policy instruments are at the disposal of the policy makers. Absent any input-output linkage, in a cooperative setting there would be no case for any trade policy intervention. With input-output linkages, absent the above mentioned first-best policies, there is a case for a second-best use of trade policy, which in this case is an import subsidy. This finding has important implications for optimal non-cooperative trade policy. The input distortion runs counter to the standard terms-of-trade considerations, thereby calling for an import tariff that is smaller in the standard setting, or for an import subsidy, even in the absence of multiple sectors.

Summing up, the general thrust of our paper is that the presence of input-output linkages in an environment of monopolistic competition establishes an economic rationale for subsidizing producers relying on material inputs. Ideally the subsidy would target the input distortion directly, but under plausible assumptions the distortion may be also addressed by means of a production subsidy. In a framework that allows for trade policy interventions, the thrust of our analysis is that the input distortion counteracts the terms-of-trade argument for an import tariff and may even call for an import subsidy.

Naturally, our analysis relies on worrying simplifications. Most importantly, we have ignored intersectoral repercussions as present in multi-sector economies, and we have assumed functional forms generating constant perceived price elasticities of demand and thus constant markups. Relaxing these simplifications are left on the research agenda.

References

[1] Baldwin, R.E. (2005), 'Heterogeneous Firms and Trade: Testable and Untestable Properties of the Melitz Model', *NBER Working Paper* No. 11471.

[2] Balistreri, E.J., R.H. Hillberry, and T.F. Rutherford (2011), 'Structural Estimation and Solution of International Trade Models with Heterogeneous Firms', *Journal of International*

Economics **83**: 95-108.

[3] Behrens, K., G. Mion, Y. Murata, J. Südekum (2018), 'Quantifying the Gap between Equilibrium and Optimum Under Monopolistic Competition', *Higher School of Economics Research Paper* No. WP BRP 185/EC/2018.

[4] Benassy, J.-P. (1996), 'Taste for Variety and Optimum Production Patterns in Monopolistic Competition', *Economics Letters* **52**: 41–47.

[5] Benassy, J.-P. (1998), 'Is There Always too Little Research in Endogenous Growth with Expanding Product Variety? *European Economic Review* **42**, 61–69.

[6] Blanchard, O.J. and N. Kiyotaki (1987), 'Monopolistic Competition and the Effects of Aggregate Demand', *American Economic Review* **77**(4): 647–666.

[7] Caliendo, L., R.C. Feenstra, J. Romalis, and A. Taylor (2017), 'Tariff Reductions, Entry, and Welfare: Theory and Evidence for the Last Two Decades', *NBER Working Paper* No. 21768, revised version.

[8] Caliendo, L. and F. Parro (2015), 'Estimates of the Trade and Welfare Effects of NAFTA', *Review of Economic Studies* **82**(1): 1-44.

[9] Costinot, A. and A. Rodríguez-Clare (2014), 'Trade Theory with Numbers: Quantifying the Consequences of Globalization', in: G. Gopinath, E. Helpman, and K. Rogoff (eds), *Handbook of International Economics* Vol 4, North-Holland.

[10] Costinot, A., A. Rodriguez-Clare, and I. Werning (2016), 'Micro to Macro: Optimal Trade Policy With Firm Heterogeneity', *NBER Working Paper* No. 21989.

[11] Demidova, S. and A. Rodríguez-Clare (2009), 'Trade Policy under Firm-level Heterogeneity in a Small Economy', *Journal of International Economics* **78**: 100-112.

[12] Dhingra, S. and J. Morrow (2019), 'Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity', *Journal of Political Economy* **127**(1): 196-232.

[13] Eaton, J. and S. Kortum (2002), 'Technology, Geography, and Trade', *Econometrica* **70**(5): 1741-1779.

[14] Epifani, P. and G. Gancia (2011), 'Trade, Markup Heterogeneity and Misallocations', *Journal of International Economics* **83**: 1-13.

[15] Ethier, W.J. (1982), 'National and International Returns to Scale in the Modern Theory of International Trade', *American Economic Review* **72**: 389?405.

[16] Felbermayr, G. B. Jung, and M. Larch (2013), 'Optimal Tariffs, Retaliation and the Welfare Loss from Tariff Wars in the Melitz Model', *Journal of International Economics* **89**(1): 13-25.

[17] Gros, D. (1987), 'A Note on the Optimal Tariff, Retaliation and the Welfare Loss from Tariff Wars in a Framework with Intra-industry Trade', *Journal of International Economics* **23**: 357-367.

[18] Jones, C. (2013), 'Misallocation, Economic Growth, and Input-Output Economics', in: D. Acemoglu, M. Arellano, and E. Dekel (eds.). Advances in Economics and Econometrics. Thenth World Congress, Volume II, Applied Economics, chapter 10. Econometric Society Monographs.

[19] Krugman, P. (1979), 'Increasing Returns, Monopolistic Competition and International Trade', *Journal of International Economics* **9**, 469–479.

[20] Krugman, P. (1980), 'Scale Economies, Product Differentiation, and the Pattern of Trade', *American Economic Review* **70**(5): 950–959.

[21] Krugman, P. and A.J. Venables (1995), 'Globalization and the Inequality of Nations', *Quarterly Journal of Economics* **CX**(4): 857-880.

[22] Melitz, M.J. (2003), 'The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity', *Econometrica* **71**(6): 1695-1725.

[23] Romer, P. (1990), 'Endogenous Technological Change', *Journal of Political Economy* **98**(5): S71-S102.

[24] Venables, A.J. (1996), 'Equilibrium Locations of Vertically Linked Industries', *International Economic Review* **37**(2): 341–359.

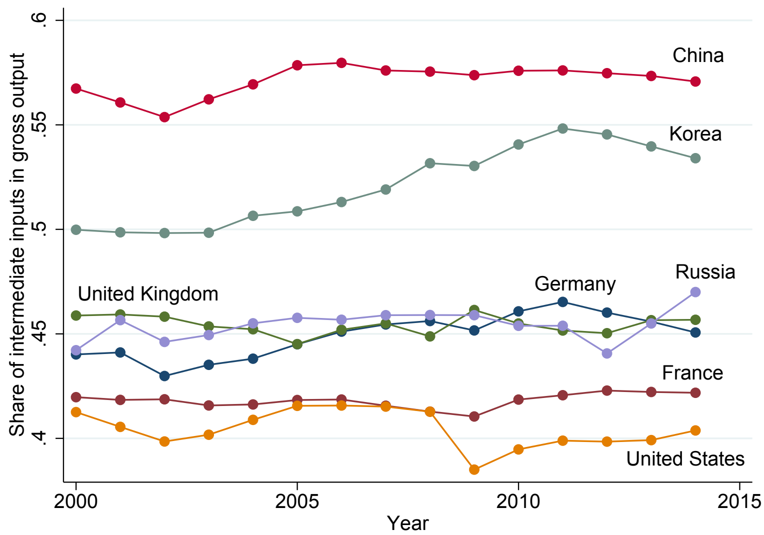
A Empirical importance of input-output linkages

We use national input-output tables for Germany, France, the UK, the US, China, and South Korea from the World input-output Database (WIOD) over the time period to compute share of intermediate inputs in gross output. Let i denote a sector. For each year, country-level shares are computed as the value-added share weighted average of sectoral intermediate input shares:

$$\sum_i \frac{\text{value added}_i}{\sum_k \text{value added}_k} \frac{\text{gross output}_i - \text{value added}_i}{\text{gross output}_i}.$$

Figure 5 shows that the shares are substantial. At the country level, they range between around. 40 percent for the US, and a bit less than 60 percent for China. Moreover, there is not much variation over time, although the period includes the year 2008, which may explain the drop for the US at that time.

Figure 5: Share of intermediate inputs in gross output



Source: World Input Database (WIOD) - National Input-Output Tables