Bavarian Graduate Program in Economics

BGPE Discussion Paper

No. 56

Trade, Productivity, and semi-endogenous Growth

Christian Bauer

June 2008

ISSN 1863-5733

Editor: Prof. Regina T. Riphahn, Ph.D.
Friedrich-Alexander-University Erlangen-Nuremberg

B G P E Discussion Paper Series

Trade, Productivity, and semi-endogenous Growth

Christian Bauer*

Abstract

We investigate the impact of incremental trade liberalization in a dynamic model of endogenous growth with heterogeneous firms and costly trade. Growth originates from horizontal specialization and the steady state productivity growth rate is positive. Innovations require costly R&D and are conducted by profit-seeking researchers. Including physical capital as a factor of production, we find that after appropriate adjustments in the production structure, previous results on the reallocation of resources and the selection of firms following trade liberalization continue to hold. We show, however, that unlike in the Melitz (2003) model, the reallocation effect does not work through increases in the factor price in production.

Keywords: Productivity, entry costs, heterogeneous firms, trade Liberalization.
JEL Classification: F12, F13, O31, O41.

*University of Regensburg, Department of Economics, Economic Theory, 93 040 Regensburg, Germany. Email: christian.bauer@wiwi.uni-r.de, Tel. 0049 941 943 2703.

1 Introduction

The relation between trade and growth remains unfinished business. On the one hand, recent empirical research convincingly argues that commonly used measures of "trade openness" are either poor measures of barriers to trade or otherwise are highly correlated with important determinants of growth (cf. Rodriguez and Rodrik, 2001). Theoretical investigations, on the other hand, highlight various specific mechanisms by which trade liberalization may affect growth and/or productivity, but this literature suffers from clear-cut results and hardly produces testable predictions. For example, trade liberalization lowers the real gross domestic product in a typical Heckscher-Ohlin model, but increases the real gross domestic product in models of monopolistic competition. Unfortunately, the key variables in competing models often correspond to different empirical measures of real income or are not observable in the data, thus making it hard to substantiate the findings. Moreover, most recent theoretical papers abstract from consumer durables and capital goods, which account for 32% and 30% of non-energy imports and 16% and 45% of non-energy exports in the U.S., respectively (Erceg, Guerrieri, and Gust, 2007).

In this paper paper, we lay out a specific environment to study how trade affects endogenous R&D in a dynamic model with heterogeneous firms and costly trade. In particular, we set up a model in which growth originates from horizontal specialization and the steady state productivity growth rate is positive. Innovations require costly R&D and are conducted by profit-seeking researchers. These features are the main difference to the canonical Melitz (2003) model.

Our model accounts for typical characteristics of both growth and trade. First, growth is semi-endogenous and thus does not display a strong scale effect. That is, the steady state productivity growth rate is exogenous, but policy makers may well exert level effects and influence the growth rate along a transition path to the steady state. Second, we account for various firm-level facts uncovered by the empirical trade literature. Most importantly, the distribution of firms' productivities is highly skewed and only the most productive firms export in equilibrium (cf. Aw, Chung, and Roberts, 2000, Bernard and Jensen, 1999, Clerides, Lach, and Tybout, 1998, Pavcnik, 2002, and Tybout, 2003, for a survey). Trade liberalization implies a reallocation of resources towards the more productive firms (cf. Melitz, 2003). Further, there is no feedback effect from exporting to a firm's productivity (Bernard and Jensen, 1999, and Bernard, Jensen and Schott, 2006). The environment laid out below, is suited to allow for both trade in final goods and trade in durables. In this paper, however, we focus on trade in intermediate goods which are produced from durable physical capital. The production of output uses specialized capital inputs and labor. Traded goods are used to produce both consumption and investment goods. Intermediate firms face endogenous fixed costs for R&D and discover production technologies with heterogenous productivities. When successful, firms enter the local product market at a cost and decide wether or not to export their goods to a foreign market. Technical barriers to trade imply that only the the most productive firms export. International trade is hampered by both variable trade costs and fixed market entry costs. Accounting for the different natures of both types of barriers to trade, we model transportation costs as capital costs and fixed trade costs as labor costs.

The reduced form of the autarky economy resembles the Jones (1995) model. Crucially, however, the productivity in R&D is not exogenous in the absence of knowledge spillovers. In this model with firm heterogeneity and market entry costs, the productivity in R&D is endogenously determined by the amount of labor necessary for market entry and the average R&D cost in the face of a minimum productivity requirement for firms.

In the open economy, we show that including trade in intermediate goods as well as production using physical capital does not alter previous findings on the reallocation of resources and the selection of firms. Similarly, modeling labor intensive technical barriers and capital intensive marginal trading costs is not essential in the baseline specification. In search of the specific mechanisms implied by the monopolistic competition heterogeneous firms models, including physical capital is an informative exercise. In Melitz (2003), trade offers additional profit opportunities only for the most productive firms. With a constant returns to scale technology, the implied market expansion effect increases the scarcity of labor, which is the only factor in production. The increase in the wage rate drives the least productive firms out of the market. In our model, the factor price for intermediate goods producing firms is independent of the exposure to trade. Furthermore, including a factor that can be accumulated potentially allows for a more pronounced impact of trade openness. The model builds on two strands of the literature, namely research on costly trade with heterogeneous firms and non-scale variety growth. We essentially include firms with heterogeneous marginal productivities and costly trade in Jones' (1995) non-scale variety growth model to account for the firm selection effect of trade openness (Bernard, Eaton, Jensen and Kortum 2003, Melitz 2003). Compared to the seminal contribution of Melitz (2003), we model endogenous entry cost and positive long-run productivity growth. Baldwin and Robert-Nicoud (2007, henceforth BRN) study these two extensions in a fully endogenous growth framework (with scale effects) and with labor as the only factor in production. They find that depending on the specification of the engine of growth, trade is likely to depress the rate of growth because with endogenous R&D, the average R&D costs are likely to increase with the necessary productivity for firms to produce profitably. Using a non-scale R&D technology, Gustafsson and Segerstrom (2007, henceforth GS) challenge this view because of the strong knowledge spillovers implicitly assumed in the BRN analysis. Using a semi-endogenous growth model, they argue that trade only has level effects. In contrast to the BRN model, more trade makes consumers better off as long as the knowledge spillovers in R&D are not too strong. Both BRN and GS focus on the effect of trade liberalization on productivity and firm selection, and thus use one factor models and perishable output. A common shortcoming is the lack of a thorough welfare analysis which is due to the complexity of the models' dynamics.

The remainder of the paper is organized as follows. We first present the closed economy model. After discussing its production structure, we characterize the autarky equilibrium. Section 3 introduces international trade. Some qualitative effects of trade liberalization are discussed in section 4. Section 5 concludes.

2 Model

Following Rivera-Batiz and Romer (1991), the world consists of two identical economies. International trade occurs only in the form of exchanges of intermediate goods. The production structure in each economy is adapted from Jones (1995), where we include heterogeneous firms and market entry costs in the spirit of Hopenhayn (1992a,b) and Melitz (2003).

2.1 Overview

Production structure. We explicitly distinguish between three sectors in each economy. The R&D sector invents blueprints for intermediate goods and conducts their market launch. Two manufacturing sectors produce intermediate goods and aggregate output, respectively.[1] Output includes consumption and investment goods.[2] There are three factors in production: labor, raw capital, and knowledge. Raw capital is the investment good, measured in terms of forgone output. The R&D technology requires labor as the only private input, and the existing stock of knowledge can have an external effect on its productivity. Aggregate output is produced from labor and a variety of imperfectly substitutable intermediate goods with additive-separable effects on output. The production of every intermediate good takes a blueprint and raw capital and is conducted by a single intermediate firm.[3] Each blueprint implies a specific level of productivity that remains constant over time.

Market entry costs. When entering the market, intermediate firms must bear a uniform entry or "beachhead" cost. Market entry is conducted using labor only, hence the entry costs take the form of a wage payment. Newly born firms make a forward looking entry decision based on their productivity. Firms which are sufficiently productive earn sufficiently high profits to cover the fixed entry cost. They therefore actually launch production in the first place and become profitable producers. Less productive firms, however, perceive that the sunk costs exceed their discounted future profits and exit right upon recognizing their productivity.

Costly trade. Each variety faces a positive demand in every country, but international trade is costly. It involves marginal trading costs as well as fixed export costs. The fixed export costs capture the additional costs a foreign company faces when selling to the local market. Importantly, country specific regulations, standards, and similar "technical" obstacles make

[1] In what follows, we use the terms "output" and "final good" interchangeably.

[2] We ignore government purchases and there will be no international trade in the final good in the open economy.

[3] We simply take firms to produce exactly one variety and equate firms with their products (i.e. good j is produced by firm j and vice versa). The boundary of intermediate firms is only essential in that we require each firm to have measure zero, so that each firm takes the price index of intermediate goods as given.

it more costly for foreign firms to enter the home market then it is for local firms.[4] The key implication of the existence of technical barriers to trade (TBTs for short) is that only the most productive firms self-select into the foreign market and earn additional profits from exporting.

Endogenous growth. Upon investing the entry costs, intermediate firms operate under monopolistic competition and earn positive profits. The prospect of these rents stimulates researchers to invent specialized inputs for the production of output. Introducing new intermediate goods continuously increases the total factor productivity (TFP) and causes growth.

Before we describe the model in greater detail, we briefly contrast the present environment with the Jones (1995) model with homogenous firms, discuss its production structure in the open economy with variable trade costs, and explain how firms with heterogeneous productivities arise from newly discovered blueprints.

2.2 Heterogeneous firms, trade, and the Jones (1995) model

Homogeneous firms, durable intermediates. The production structure of the Jones (1995) model is taken from Romer (1990). In Romer (1990), the capital stock comprises a continuum of durable capital goods, which imperfectly substitute in the production of output, with additively separable effects.[5] The capital goods are assembled by intermediate firms. Using $k(j)$ units of the investment good, firm j assembles $x(j) = k(j)$ units of the specialized capital good j. The investment good, "raw capital", is produced from labor and existing durable goods. It is convenient and common practice to assume identical production technologies for the consumption good and the investment good so that the output from both sectors can be summarized as aggregate output which can either be used for investment or for consumption. Romer (1990) already noted that the one-to-one production of intermediate goods from raw capital is merely assumed to keep the model simple. Similarly, uniform production technologies across interme-

[4]See Baldwin (2000) for an illustrative introduction to technical barriers to trade.

[5]Breaking up the capital stock in a continuum of imperfectly substitutable goods allows for positive market rents, which are necessary to cover the innovation costs when production technologies are not strictly convex (see, among others, Romer, 1990).

diate firms are typically used only for analytical convenience.

Heterogenous firms. In this research, intermediate firms are heterogeneous with respect to their productivity. We thereby incrementally extend two workhorse models. First, relative to the Jones (1995) model, the average "efficiency" of intermediate firms contributes as a second, "vertical" dimension of productivity to the level of TFP.[6] The range, and along with it the average of firms' productivities in production, is endogenously determined by the degree of trade openness as measured by trade costs. In contrast to growth models with both horizontal and vertical innovations, only the number of varieties increases continuously over time (R&D with heterogeneous firms is addressed in detail in the next but one paragraph). Second, relative to the existing literature on growth and trade with heterogeneous firms, intermediate goods are not only used for consumption, but also for investment. This extension opens up the possibility of a more pronounced impact of trade. Accounting for the accumulation of physical capital, we further add a second factor in production.

Marginal trade costs and the allocation of capital. The presence of marginal trade costs requires a careful modeling of the spatial allocation of physical capital. The production structure of the Jones (1995) model in principle allows two equitable interpretations. The first, classical interpretation (used by Romer, 1990 and Jones, 1995) is that intermediate goods are durable inputs in the production of output. Intermediate good producing firms assemble the durables from raw capital and pass the processed capital on to output producing firms. In this case, capital accumulates at the location of the final good production. In the second interpretation, raw capital is a durable good in the production of intermediate goods. Intermediate firms accumulate physical capital to produce perishable inputs for the production of aggregate output. In this case, the capital stock is located at the origin of the intermediate good production.

No trade in durable commodities. In the closed economy, both interpretations are equivalent. As long as there are no variable transportation cost, both interpretations are equiv-

[6]Li (2000), Young (1998), and Kornprobst (2008, Ch. 9) present models with two R&D sectors and both horizontal and vertical innovations. Sorger (2007) considers quality improving horizontal innovations in a one-sector R&D model, where researchers can influence the quality of their innovations at the cost of a reduced quantity of innovations.

alent in the open economy as well. To simplify matters, in what follows, we focus on perishable inputs in the production of durable investment and consumption goods (we stick to the second interpretation above). Since we also rule out trade in aggregate output, there is no accumulation of physical capital by imports.[7] From an empirical point of view, neglecting trade in durable/capital goods appears as a severe shortcut. Erceg, Guerrieri, and Gust (2007) find for the U.S. that consumer durables and capital goods amount to 32% and 30% of non-energy imports, and 16% and 45% of non-energy exports, respectively. In their data, consumer non-durables represent about one-fourth of non-energy imports and exports. The remainder is non-energy industrial supplies used in the production of durables.

Variety expanding R&D and heterogeneous firms. The discovery of blueprints for new intermediate goods is at the heart of our model of growth and trade. A crucial question is how labor and knowledge are transformed into blueprints with heterogeneous productivities. We adapt the modeling in BRN, but use a non-scale technology like GS. Following Melitz (2003), the productivity types of blueprints are drawn from a given stationary distribution. The resources necessary to produce a sufficiently valuable blueprint however are endogenously determined.

Stochastic productivity draws. While researchers can be certain about finding a new blueprint, its inherent productivity is random. Every research attempt is a costly draw. Due to the entry costs, only blueprints with a sufficiently high productivity (and hence a sufficiently high market value) sell at a positive price. For the sake of clarity, we formally treat R&D and manufacturing as performed in separate sectors. As regards content, we may equivalently combine the two activities for a given variety in "the firm". With a slight abuse of terms, we then also call costly developed blueprints which do not make it into the product market "firms". This gives us a theoretical counterpart to those very low productivity type firms for which the empirical trade literature has identified a high death rate. In the model, these "firms" exit immediately upon recognizing their productivity.

Costly aggregate productivity gains. One of the contributions of BRN is to incorporate the idea that increasing the productivity of innovations is costly, in the sense that R&D (c.p.

[7]The "trade in intermediate goods only" approach follows Rivera-Batiz and Romer (1991).

and on average) requires more resources if its outcome is to be more productive. In modeling this notion, BRN look at R&D from an aggregate point of view and consider the average costs associated with the discovery of a marketable blueprint. A potential drawback of this "aggregate R&D" approach is the lack of intentional investments in more productive capital goods. In fact, individual researchers cannot influence the productivities of their innovations. From the individual researcher's perspective, conditional on being usable, high productivity type blueprints are "lucky draws" and as such, they come for free: every draw is equally costly. As will be discussed in more detail in Section 2.5, free entry into R&D does not remove the windfall gains associated with high productivities because researchers must break even across usable and unusable innovations in expectations.

A productivity frontier in R&D. As a final remark, note that there is a close analogy between the "aggregate R&D" approach and a productivity-quantity frontier in R&D. That is, an increase in the quality of products will c.p. come at the cost of fewer innovations.[8] Increasing the minimum productivity requirement (again c.p.) forces researchers to move along the technologically given productivity-quantity frontier towards more productive blueprints and fewer innovations. Trade liberalization, as measured by a decrease in the foreign market entry costs, actually raises the minimum productivity requirement, thereby increasing the average productivity of intermediate firms. This productivity gain however is not "manna from heaven" but takes costly resources and implies that the set of intermediate goods at least temporarily expands at a lower rate. Via this channel, the exposure to trade has the potential to slow down productivity gains from specialization. Hence, trade liberalization may at least temporarily depress growth and at the same time have ambiguous effects on TFP.

To begin with, we show how endogenous horizontal innovation and TFP is affected by a minimum productivity requirement in autarky. We then turn to the open economy with international trade in section 3.1.

[8] Sorger (2007) explicitly includes such a frontier in R&D in a closed economy, free entry model of variety growth. In his model, researchers choose the quality of their innovations optimally, recognizing that higher qualities imply fewer R&D output (cf. footnote 6).

2.3 Autarky

The economy is characterized by preferences, endowments, technologies, and a specific institutional environment. As in Romer (1990), Jones (1995), or Rivera-Batiz and Romer (1991), the specific environment laid out below allows a concise exposition and is only one example of an environment that supports the decentralization. The model is set in continuous time and final output is used as the numéraire. We omit the time argument, t, wherever it is not confusing, and occasionally abbreviate variables in the argument of functions by a centered dot ("$\cdot$").

2.3.1 Households

The economy is populated by a continuum of mass one of identical households. Every household consists of L homogenous members, who inelastically supply one unit of labor each (there is no disutility from work). The population grows at an exogenously given, constant rate $\dot{L}/L \equiv n \geq 0$, and $L(0) > 0$.[9] The households are infinitely-lived Barrovian (1974) dynasties, where each generation cares about the well-being of all its future offsprings. Every household member consumes an equal amount c of aggregate output Y. The consumption behavior is therefore appropriately summarized by the optimal decision of one household. Preferences are given by a standard intertemporal utility function with constant intertemporal elasticity of substitution in consumption equal to $1/\sigma$ (≥ 0):[10]

$$U = \int_0^\infty L(t)e^{-\rho t}u(c(t))dt, \quad u(c(t)) = \frac{c(t)^{1-\sigma} - 1}{1 - \sigma}.$$

ρ (> 0) is the subjective discount rate.

Every household earns income from working and returns on assets and purchases consumption goods and assets. The flow budget constraint is $\dot{\zeta} = wL + r\zeta - cL$, where wL and cL denote the household's labor income and consumption, respectively, and $r\zeta$ is the return on asset holdings ζ at interest rate r. Assets comprise ownership claims on physical and financial

[9] Arnold (1998) replaces population growth with human capital accumulation in a Grossman-Helpman (1991, Ch. 3) framework (without physical capital) and thereby shows explicitly that L can be interpreted more broadly as the effective labor force.

[10] The elasticity of marginal utility is also constant and equals $-\sigma$.

capital (loans and debts between households cancel in the representative households' budget constraint). Subsequent assumptions on the observability of firm types and the capital market ensure that physical capital and all types of equity are perfect substitutes as vehicles of savings. They all pay a common rate of return r.

Ponzi-games, where some households borrow infinitely to "repay" consumption loans (and in fact never actually repay their credit), are ruled out by a borrowing constraint imposed in the capital market. Bankers will not lend out more than the present value of a household's income. Hence the present value of consumption expenditures is bounded above by the present value of income. As usual, the appropriate condition is that the present value of assets is asymptotically non-negative, $\lim_{t\to\infty}\left\{\zeta(t)\exp\left[-\int_0^t r(s)ds+nt\right]\right\}\geq 0$.[11]

2.3.2 Technology in manufacturing

Output. Aggregate output Y is produced using a set of measure A of vertically differentiated intermediate goods j in quantities $x(j)$ and labor L_Y:

$$Y = L_Y^{1-\alpha}\int_0^A x(j)^\alpha dj, \quad 0<\alpha<1. \tag{1}$$

Output is manufactured by a large number of identical firms (the number of firms is indeterminate because of constant returns to scale for a given level of A).[12] Labor and intermediate goods are complements ($\partial^2 Y/(\partial x \partial L_Y)>0$). The elasticity of substitution between any pair of intermediates is $(1<)\ \epsilon\equiv 1/(1-\alpha)\ (<\infty)$. Given the parameter restriction implicit in (1), the intermediate goods have an additively separable effect on output ($\partial^2 Y/[\partial x(j)\partial x(j')]=0$).[13] As usual, the parameter α jointly determines the returns to horizontal specialization in the production of output, the elasticity of substitution between intermediate goods (which indicates

[11]Non-negativity constraints on consumption can be ignored as the instantaneous utility function $u(c)$ satisfies $u'(c)\to\infty$ as $c\to 0$.

[12]The production function in (1) of course displays increasing returns in L_Y, all $x(j)$, and A jointly.

[13]Alvarez-Pelaez and Groth (2005) introduce a more general production function $Y = A^\gamma X^\beta L_Y^{1-\beta}$, $X = A\left[\frac{1}{A}\int_0^A x(j)^\alpha dj\right]^{\frac{1}{\alpha}}$ where intermediates can be substitutes ($\alpha>\beta$) or complements ($\alpha<\beta$). We implicitly impose $\gamma = 1-\beta$ and $\alpha=\beta$ for simplicity.

the degree of market power of intermediate producers), the price elasticity of demand, and also pins down constant shares of factor incomes in equilibrium.[14]

Intermediates. Every intermediate good is produced from raw capital by an intermediate firm that exclusively owns its blueprint. Each blueprint implies a constant level of productivity in production which carries over to its producer. The firm-level differences in productivities are captured by heterogenous per unit input coefficients $b(j)$:

$$x(j) = \frac{k(j)}{b(j)}, \quad b(j) \in (0, b_0]. \tag{2}$$

More productive firms, i.e. firms with low $b(j)$, require less raw capital $k(j)$ to produce one unit of their intermediate good. Unlike in the original Romer model (1990), we treat raw capital as a durable good in the production of perishable intermediate goods. The production and export of the intermediates implies a permanent flow of production and transport costs and simplifies the open economy model in Section 3.

2.3.3 Technology in R&D

The presence of entry costs implies that forward looking, profit-driven firms only launch production with blueprints that yield a positive operating profit. Firms' profits are obviously increasing in productivity, which implies that the lowest productivity-type blueprints will be discarded due to the entry costs. If this minimum productivity requirement is binding, the number of intermediate goods (A) is lower than the total number of discovered blueprints (B). In Romer (1990) and Jones (1995), there are no barriers to entry and every discovered blueprint is used to produce a new variety ($A = B$). To tackle this issue, we may think of the R&D technology conceptually as involving two parts. "Research" comprises the process of discovering a previously unknown blueprint. "Development" involves the productivity in production inherent in each blueprint. We consider both parts in turn.

[14]It is possible to disentangle the elasticity of output with respect to (horizontal) specialization and the substitutability of capital goods, see Benassy (1998). Alvarez-Pelaez and Groth (2005) also disentangle the degree of substitutability from the capital share, see footnote 13.

Discovery of blueprints. Researchers deterministically invent new blueprints $\dot{B}$ using Jones' (1995) R&D technology:

$$\dot{B} = \frac{L_B A^{1-\chi}}{F_B}, \quad \chi > 0, \; F_B > 0. \tag{3}$$

L_B is the number of people searching for new blueprints, and F_B inversely measures their productivity. Following common practice in endogenous growth theory, innovation displays constant returns to scale in its only private input, labor. Previous research efforts can have external effects on the magnitude of labor required for innovation, and we follow BRN in choosing the existing number of intermediate goods (A) to represent the relevant knowledge stock.[15] The exponent $1 - \chi$ accounts for the strength and the sign of the knowledge spillovers. Researchers may either "stand on the shoulders of giants" and benefit from past innovations ($\chi < 1$) or face the "fishing out" of ideas ($\chi > 1$). If $\chi = 1$, there are no spillovers. In this case,

$$\frac{\dot{B}}{B} = \frac{L_B}{BF_B}$$

so that the growth rate of B declines if B increases for a given L_B. It then takes positive growth of the labor input to maintain positive long-run growth. At $t = 0$, the economy is endowed with a mass $B(0) = B_0$ of blueprints with distribution $G(b)$.

Jones' (1995) R&D technology is intended to eliminate the strong scale effect, i.e. the dependence of the productivity growth rate on the level of labor engaged in R&D in the long run. In doing so, his specification "exogenizes" long-run growth. Suppose $A = B$ and $\dot{L}_B/L_B$ (as is the case along a balanced growth path in Jones' model). Then, a constant growth rate of the number of blueprints requires $n - \chi\dot{B}/B = 0$, or $\dot{B}/B = n/\chi$.[16] Thus, growth is semi-endogenous (in that the long run growth rate cannot be influenced by policy) and trade liberalization can "only" exert level effects.

Having described the discovery process, we now turn to the productivity in production

[15] Without intentional investments in qualities, B seems equally appropriate as A to represent past innovation efforts.

[16] In the aforementioned Sorger (2007) model with intentional investment in quality, growth does not display a strong scale effect. In his model, policy makers can affect the growth rate if they are able to design quality contingent subsidies (see also Howitt, 1998).

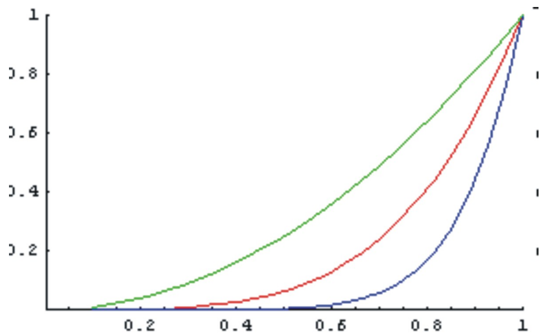

Figure 1: The Pareto distribution of input coefficients.

that comes along with each blueprint.

Stochastic assignment of productivities. The level of productivity is indicated by variety-specific input coefficients, which are randomly assigned to each blueprint and revealed after the R&D investment is made (i.e., at the time a blueprint is discovered). The input coefficients are drawn from a distribution which has many low productivity types, fewer intermediate productivity types, and only a few types of very high productivity. To be specific, the input coefficients are drawn from the "mirrored" Pareto distribution

$$G(b) = (b/b_0)^{\theta}, \; b \in [0, b_0], \tag{4}$$

where the parameters b_0 (> 0) and

$$\theta > \max\{\epsilon - 1, 1\}$$

govern the width of the support and the shape of the cumulative distribution function, respectively. Figure 2.3.3 depicts the distribution of input coefficients for $\theta = 2$ (blue), $\theta = 4$ (red), and $\theta = 8$ (green) with $b_0 = 1$. Imposing a lower bound on θ serves two purposes. First, as will become clear below, $\theta > \epsilon - 1$ ensures that the input coefficient of the least productive firm is strictly positive (so that there is a non-degenerated distribution of firms). Second, it preserves the intended skewness towards low productivity types in case of $\alpha < 0.5$ (in this case, $\theta > \epsilon - 1$ does not imply $\theta > 1$).[17] θ measures the steepness or "dispersion" of the distribution and can

[17]Both BRN and GS do not impose the second parameter restriction which, however, is only important for the interpretation.

therefore be interpreted as the inherent likelihood (or "difficulty") of inventing high productivity types. Increasing θ gives first-order stochastically dominated distributions, i.e. distributions that are more skewed towards high input coefficients ($\theta = 0$ is the uniform distribution and $\theta \to \infty$ yields a degenerate distribution at b_0, in which case $G(b) \to 0$ for all $b < b_0$).[18]

From blueprints to firms. The distribution underlying the productivity types of newly discovered blueprints directly translates into the productivity distribution of firms. This is because the Pareto distribution has the property of scale invariance: truncating a Pareto distribution yields another Pareto distribution with the same shape parameter.[19] As an example, suppose that the cumulative distribution function $G(b)$ is truncated at some minimum productivity $1/b_{trunc}$. The resulting distribution of input coefficients is

$$G(b|b \leq b_{trunc}) = \frac{G(b)}{G(b_{trunc})} = \frac{\left(\frac{b}{b_0}\right)^{\theta}}{\left(\frac{b_{trunc}}{b_0}\right)^{\theta}} = \left(\frac{b}{b_{trunc}}\right)^{\theta}$$

for $b \in [0, b_{trunc}]$. Thus, when some blueprints are not used due to the minimum productivity requirement, the distribution of firms productivities still remains Pareto, and θ equivalently reflects the dispersion in the truncated distribution (the support simply shrinks from $[0, b_0]$ to $[0, b_{trunc}]$). Given the shape of the underlying productivity distribution, the distribution of firms' productivities matches the empirical regularity that the proportion of less productive firms is large.

Justifying the Pareto distribution. Like BRN and GS, we specify a functional form to obtain a closed form solution. The Pareto distribution is attractive for two reasons. Firstly, it receives strong empirical support when it comes to matching the observable distribution of productivities, see e.g. Cabral and Mata (2003) and Corcos, Del Gatto, Mion, and Ottaviano (2007). Secondly, as pointed out above, it allows a tractable analytical exposition of the distribution of firm types because truncating a Pareto distribution yields another Pareto distribution with the same shape parameter (it is scale invariant).

[18] The expected value and variance are $E(b) = \int_0^{b_0} b dG(b) = \int_0^{b_0} \frac{\theta}{b_0^{\theta}} b^{\theta} db = \frac{\theta}{b_0^{\theta}} \left[\frac{b^{\theta+1}}{1+\theta}\right]_0^{b_0} = \frac{\theta}{1+\theta} b_0$ and $Var(b) = E(b^2) - [E(b)]^2 = \int_0^{b_0} b^2 dG(b) - \left(\frac{\theta}{\theta+2} b_0\right)^2 = \frac{\theta}{(2+\theta)(1+\theta)^2} b_0^2$ (which is decreasing in θ).

[19] More generally, the Pareto distribution belongs to the class of power law distributions, which are characterized by the scale invariance property (θ is then consistently called the scaling parameter).

2.3.4 Markets

The markets for labor, the final good, and financial capital are all perfectly competitive. Producers of capital goods hold infinitely-lived, fully enforced patents. All markets clear. Ownership claims on physical capital and financial wealth are perfect substitutes and pay the same rate of return, r.

Fundamental evaluation. Once a firm's input coefficient is revealed (upon discovery of its blueprint), it immediately becomes common knowledge. We denote by $\pi(j)$ the instantaneous profits of firm j and let

$$v(j) \equiv \int_t^\infty e^{-\bar{r}(s-t)}\pi(j)ds, \tag{5}$$

where $\bar{r} \equiv \int_t^s r(\varsigma)d\varsigma$ is the cumulative interest rate up to time $s \geq t$. In the absence of bubbles, and due to the sunk nature of both innovation and entry costs, $v(j)$ is the market value of firm j (with input coefficient $b(j)$). Differentiating (5) with respect to time t reveals that, given the definition of $v(j)$ as fundamental value, the returns from investing in any productivity-type of firm, i.e. the dividend payments plus capital gains, have to equal the common return on either asset:

$$\pi(j) + \dot{v}(j) = rv(j) \quad \forall j \in [0, A]\,. \tag{6}$$

Market clearing. Labor market clearing requires that the sum of labor in innovation, market entry, and production is equal to the labor force,

$$L = L_B + L_E + L_Y. \tag{7}$$

We further denote by $L_A \equiv L_B + L_E$ the total labor force engaged in the process of R&D and market entry, which we henceforth refer to as R&E (a mnemonic for R&D plus entry).

The stock of raw capital is

$$K \equiv \int_0^A k(j)dj. \tag{8}$$

Capital does not depreciate. In Jones (1995), where $b(j) = b = 1$, the sum of intermediate goods equals the amount of accumulated forgone consumption, i.e., the stock of raw capital. Here, with heterogeneously productive firms, the sum of intermediate outputs is proportional to the stock of raw capital and the factor of proportionality equals the output weighted average input coefficient.[20] From (2) and (8)

$$K = \int_0^A b(j)x(j)dj. \tag{9}$$

If intermediate firms become more productive on average, an increased amount of intermediate goods can be obtained from forgoing a given amount of consumption.[21]

Finally, the resource constraint defined over economy-wide aggregates is

$$Y = cL + \dot{K}. \tag{10}$$

2.3.5 Market Entry

Launching the production of a newly discovered capital good is equally costly to all entrants. To keep the analytical exposition simple, we follow BRN and assume identical production functions (and thereby "factor intensities") in R&D and the conduct of entry. The productivity in the entry process thereby indicates the markets' "openness". Strictly speaking, the entrant is required to hire $A^{\chi-1}F_L$ workers and pay the associated wage bill $wA^{\chi-1}F_L$. F_L measures the strength of the barriers to entry.[22] To ensure that the input coefficient of the least productive firm in equilibrium is strictly smaller than the upper bound of the underlying distribution, b_0 (i.e. that the minimum productivity requirement introduced by the entry cost is binding in equilibrium), we impose a lower bound on F_L:

$$F_B < (\phi - 1)b_0^\theta F_L. \tag{PA1}$$

At any point in time, the economy-wide amount of labor devoted to preparing entry is

$$L_E = \dot{A}A^{\chi-1}F_L. \tag{11}$$

[20] Given mark-up pricing in the intermediate good sector, the output weighted average productivity is closely related to the CES price index. In his one-factor zero-growth model, Melitz (2003, footnote 9) uses such a output-weighted average to measure overall productivity.

[21] As pointed out in the model introduction, assuming that capital can be accumulated as forgone output implies that raw capital is produced with the same technology as the final good. "Forgone consumption" in the above interpretation is thus not actually produced in the first place, but the respective resources are used to produce, i.e. accumulate, raw capital instead.

[22] In BRN and GS, the interpretation of the innovation and entry process is that researchers have to accumulate F_B units of knowledge for inventing a new blueprint and F_L units of knowledge to cope with market entry.

Since all productivities are immediately revealed and become common knowledge when the blueprint is discovered, the entry decision involves no uncertainty.[23]

Justifying the entry specification. Four remarks on the specification of entry costs are in order. First, the scaling of entry costs by $A^{\chi-1}$ makes a balanced growth equilibrium with a constant ratio of entry costs and the market value of a new capital good (which, by construction, lies between zero and one) possible. Without resorting to (completely) arbitrary scaling factors, we could alternatively use $F_L K/A$ or $F_L Y/A$ (and include the use of resources in the respective market clearing/resource condition). Second, identical production functions in R&D and entry turn out to be particularly convenient because they allow a manageable analytical treatment of the free entry into R&D condition. Third, exploiting the block-recursive structure of the Jones (1995) model, identical "factor intensities" in R&D and entry allow simple aggregations of both processes. Fourth, in the open economy, trade is restricted by marginal costs and TBTs. Modeling variable trade costs as iceberg costs implies that they are capital costs. With respect to the nature of TBTs, we assume that overcoming technical obstacles is by far more labor intensive, and take the extreme standpoint that fixed barriers to trade imply only labor costs. Equilibrium

Having described the environment, we now derive optimality conditions, define the equilibrium, and aggregate over the different types of firms. The subsequent section then characterizes the equilibrium balanced growth path.

2.4 Optimality conditions

Households and firms maximize their utility and profits, respectively. We consider their decisions in turn.

[23] This timing structure emphasizes the importance of entry cost. If researchers individually knew the productivity of their future innovations, the sunk innovation cost would obviously be sufficient to prevent low productivity types from being invented in the first place.

2.4.1 Households

Optimal behavior of households boils down to choosing a path for consumption. Given a measure $B \geq B_0$ of firms, households are able to pool the risk of investing firms whose type is a priori unknown. Hence, optimal consumption is not affected by the actually prevailing productivity distribution of firms in a household's portfolio or in the economy. Maximizing intertemporal utility subject to the flow budget constraint and the no-Ponzi game condition (or, equivalently, to an intertemporal budget constraint that limits the present value of consumption spending to the present value of total income) yields the well-known Euler equation

$$\frac{\dot{c}}{c} = \frac{r - \rho - n}{\sigma} \tag{12}$$

and a transversality condition.[24] As usual, the Euler equation gives the rate of consumption growth that optimally relates the subjective discount rate (including household growth) and the market interest rate.

---

[24] If households maximize utility in per capita terms, the present value Hamiltonian is

$$H = e^{-\rho t} u(c) + \lambda (wL + r\zeta - cL),$$

where λ represents the shadow price of wealth. H is concave in c and ζ, so that the following first-order conditions are sufficient for optimality:

$$\begin{aligned} \frac{\partial H}{\partial c} &= e^{-\rho t} c^{-\sigma} - \lambda L \overset{!}{=} 0, \\ \frac{\partial H}{\partial \zeta} &= r\lambda \overset{!}{=} -\dot{\lambda}, \\ \lim_{t \to \infty} \zeta\lambda &= 0. \end{aligned}$$

Inserting $\lambda = e^{-\rho t} c^{-\sigma} / L$ from the first condition and its time derivative,

$$\begin{aligned} \dot{\lambda} &= \frac{L \left(-\rho e^{-\rho t} c^{-\sigma} - \sigma e^{-\rho t} c^{-\sigma - 1} \dot{c} \right) - e^{-\rho t} c^{-\sigma} \dot{L}}{L^2} \\ &= \frac{e^{-\rho t} c^{-\sigma}}{L} \left(-\rho - n - \sigma \frac{\dot{c}}{c} \right), \end{aligned}$$

in the second optimality condition yields (12). Substituting $\lambda = e^{-\rho t} c^{-\sigma} / L$ and $u'(c) = c^{-\sigma}$ in the third optimality condition, the transversality condition requires that households must not get any utility out assets as $t \to \infty$,

$$\lim_{t \to \infty} \frac{\zeta e^{-\rho t} c^{-\sigma}}{L} = \frac{e^{-\rho t} u'(c)}{L} = 0.$$

2.4.2 Firms

Profit maximization and competition in the output producing sector imply that the aggregate demand for production workers L_Y and intermediate goods $x(j)$, $j \in [0, A]$, satisfy

$$\begin{aligned} L_Y &= \frac{(1-\alpha)Y}{w}, & (13) \\ x(j) &= \left[\frac{\alpha}{p(j)}\right]^{\epsilon} L_Y. & (14) \end{aligned}$$

As mentioned earlier, the price elasticity of demand is

$$\frac{\partial x\left(j\right) p\left(j\right)}{\partial p\left(j\right) x\left(j\right)} = -\epsilon.$$

Given the demand function in (14), every intermediate goods producer producing firm maximizes its profit $\pi(j)$ by charging a price equal to a constant mark-up over the firm-specific marginal cost (irrespective of the time of invention):

$$p(j) = \frac{rb(j)}{\alpha}, \quad \forall j \in [0, A]. \tag{15}$$

Using (15) in (14), the equilibrium demand and revenues $R\left(j\right) \equiv p\left(j\right) x\left(j\right)$ of firm j with input coefficient $b(j)$ are

$$\begin{aligned} x(j) &= \alpha^{2\epsilon} \left[rb(j)\right]^{-\epsilon} L_Y, & (16) \\ R(j) &= \alpha^{2\epsilon-1} \left[rb(j)\right]^{1-\epsilon} L_Y, \quad \forall j \in [0, A]. & (17) \end{aligned}$$

From (15), profits amount to

$$\pi(j) = (1-\alpha)R(j), \quad \forall j \in [0, A]. \tag{18}$$

Obviously, profits are increasing in productivity $1/b$. From (18),

$$\frac{\partial \pi\left(b, \cdot\right)}{\partial\left(\frac{1}{b}\right)} = \left(1-\alpha\right) \alpha^{2\epsilon-1} \left(\epsilon - 1\right) \left(\frac{1}{b}\right)^{\epsilon-2} \left(\frac{1}{r}\right)^{\epsilon-1} L_Y,$$

which implies that profits are convex (concave) in productivity if $\epsilon > 2$ ($\epsilon < 2$), i.e. $\alpha > (<) 1/2$.

Gains from increasing degrees of specialization with imperfectly substitutable intermediate goods limits a complete allocation of resources towards the most productive firms. In particular, the market entry costs are the necessary ingredient to prevent the least productive firms from operating: There is always a positive demand for any variety as long as any output is produced ($L_Y > 0$), and mark-up pricing guarantees positive operating profits for firms of all productivity types. In the absence of barriers to entry ($F_L = 0$), all firms launch production, $b_L^* = b_0$, so that $A = B$.

No durable goods monopoly problem. Note that our interpretation of the production structure with durable goods in the intermediate rather than the final good sector naturally avoids the usual "durable goods monopoly problem". When monopolists actually sell durable goods, tomorrow's demand is a close substitute to today's demand, and firms with market power account for the fact that today's sales come at the expense of tomorrow's sales. Tirole (1988, section 1.5) shows that monopolists then have an incentive to increase today's quantities at the expense of tomorrow's demand and do so in the absence of commitment to output quantities. Romer (1990) points out that in his model environment, selling durable goods to the final good sector potentially results in a more complicated pricing problem than the "static" program stated above. To avoid this complication, Romer (1990, in a closed economy) and Rivera-Batiz and Romer (1991, in an open economy) formally assume that the durable goods are rented. In our interpretation, the problem is resolved since there is no monopolistic supplier of the investment good.

From goods to productivities. In this environment, the intermediate firms' prices, quantities, profits, and firm values differ only due to heterogeneous productivities. As of this point, it is thus reasonable to drop the firm index j and phrase the equilibrium expressions in terms of productivity types b, i.e. from (15) and (16),

$$p(b, \cdot) = \frac{rb}{\alpha}, \quad x(b, \cdot) = \alpha^{2\epsilon}(rb)^{-\epsilon} L_Y, \tag{19}$$

and from (18) and the definition of the firm value in (5),

$$\pi(b, \cdot) = (1 - \alpha)\alpha^{2\epsilon-1}(rb)^{1-\epsilon} L_Y, \quad v(b, \cdot) = \int_t^{\infty} e^{-\bar{r}(s-t)}\pi(b, \cdot)ds. \tag{20}$$

Similarly, the time derivatives of the firm values in (6) simplify to

$$rv(b, \cdot) = \pi(b, \cdot) + \dot{v}(b, \cdot). \tag{21}$$

Understanding firm heterogeneity. To improve our understanding of firm heterogeneity in this production environment, consider a firm with input coefficient b that is more efficient than another firm with input coefficient $b' \geq b(j)$. From (19), we find that relative output is

$$\frac{x(b,\cdot)}{x(b',\cdot)} = \left(\frac{b}{b'}\right)^{-\epsilon} = \left(\frac{b'}{b}\right)^{\epsilon} \quad (\geq 1).$$

Similarly, the relative input requirement in the production is

$$\frac{bx(b,\cdot)}{b'x(b',\cdot)} = \left(\frac{b}{b'}\right)^{1-\epsilon} = \left(\frac{b'}{b}\right)^{\epsilon-1} \quad (\geq 1).$$

The relative output and input quantities are thus independent of endogenous variables, and the only parameter besides the productivities themselves that has an impact at all is α. Finally,

$$\frac{v(b,\cdot)}{v(b',\cdot)} = \frac{\pi(b,\cdot)}{\pi(b',\cdot)} = \frac{R(b,\cdot)}{R(b',\cdot)} = \left(\frac{b}{b'}\right)^{1-\epsilon} = \left(\frac{b'}{b}\right)^{\epsilon-1} \quad (\geq 1).$$

The second equality holds because of our assumptions on the fundamental capital market evaluation above. Since the input coefficients are constant over time, the profits of firms of all productivity types, and hence their market values, grow at equal rates:

$$\frac{\dot{v}\,(b,\cdot)}{v(b,\cdot)} = r - \frac{\pi(b,\cdot)}{v(b,\cdot)} = r - \frac{\pi(b,\cdot)}{\int_t^\infty e^{-\bar{r}(s-t)}\pi(b,\cdot)ds}, \tag{22}$$

and b cancels from the last term (because it can be pulled out of the integral). Hence, $\hat{v}\,(b,\cdot) = \hat{v}\,(j) = \hat{v}$ so that the dividend ratio is identical across firms of all productivity types. In equilibrium, firms with a higher productivity sell higher quantities, demand more raw capital (as the lower input coefficient is offset by the rise in total demand), receive higher profits, and have a higher market value. An increase in α amplifies the differences. Figure 2 depicts a firm's profit and its market value as a function of its productivity for $\epsilon < 2$.[25] We summarize these findings in

Result 1 (Productivity and firm size)**.** *In equilibrium, more efficient firms are larger: they produce more output and use more raw capital than less efficient firms. Profits and firm values are increase and concave (convex) in the firm's productivity if* $\epsilon < (>)2$.

[25] Differentiating equilibrium profits with respect to b yields $\frac{\partial \pi(b,\cdot)}{\partial b} = \epsilon(\epsilon-1)\frac{\pi(b,\cdot)}{b^2} > 0$. This immediately gives a firm value function that is of the same shape, since the dividend ratio is the same for all productivity type firms which implies that the ratio of the slopes of v and π is identical across productivity types.

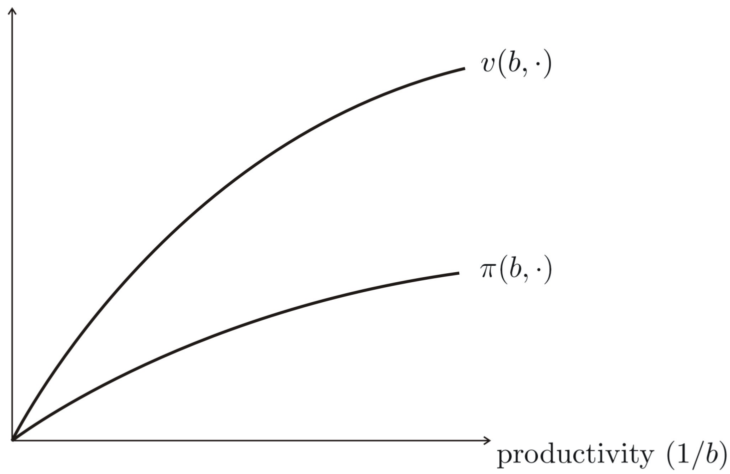

Figure 2: Firm's profits and value as a function of productivity ($\epsilon < 2$).

Obviously, higher input prices ($r \uparrow$), and less demand from the final good sector ($L_Y \downarrow$) c.p. imply smaller profits. Clearly also, the profits of more efficient firms react stronger to such changes in absolute terms (here exemplarily for r):

$$\frac{\partial \pi(j)/\partial r}{\partial \pi(j')/\partial r} = \left(\frac{b'}{b}\right)^{\epsilon-1} \quad (> 1).$$

Profits and α. The relation between firms' profits and the parameter α deserves a short comment. As pointed out above, changing α has multiple implications, and it also captures opposing effects on intermediate firms' profits. On the one hand, like in the canonical trade models with love of variety preferences, a low degree of substitutability between the differentiated final good inputs (a low α) allows the monopolists to charge a high mark-up $1/\alpha$, and (as demand is inelastic) earn high revenues and high profits. On the other hand, α also measures the capital share in the production of final output. Hence, a small α also presumes less demand for capital goods from producers of output goods. Using standard parameters, the latter effect prevails and profits are increasing in α.[26]

[26] In fact, the sign of the net effect actually depends on the size of the input coefficient. From (20),

$$\ln \pi\left(b,\cdot\right) = \ln\left(1-\alpha\right) + \left(2\epsilon-1\right)\ln\alpha + \left(1-\epsilon\right)\ln\left(rb\right) + \ln L_Y,$$

and hence

$$\frac{\partial \ln \pi\left(b,\cdot\right)}{\partial \alpha} = \frac{1}{\alpha-1} + \frac{2\epsilon-1}{\alpha} + 2\frac{\partial \epsilon}{\partial \alpha}\ln\alpha - \frac{\partial \epsilon}{\partial \alpha}\ln\left(rb\right).$$

2.4.3 Entry

Let us return to the entry decision of the firm. The imposed upper bound on F_B restricts the analysis to the case where F_L (or b_0) is sufficiently "large" so that the entry costs exceed the market value of the least productive firms (i.e. $v(b_0, \cdot) < wA^{1-\chi}F_L$ holds in equilibrium by assumption). Thus, only sufficiently productive firms are willing to bear the entry cost. Given market prices, the cutoff productivity associated with profitable entry, $1/b_L$, is determined by

$$v(b_L, \cdot) \equiv wA^{\chi-1}F_L. \tag{23}$$

Equation (23) is illustrated in Figure 3. Firms with a productivity below $1/b_L$ will not incur the entry costs and "die" instantaneously. More productive firms incur the costs and launch production. Due to the scale invariant nature of the Pareto distribution, whereby truncating the distribution maintains both the type of the distribution and its shape parameter, all information about the equilibrium distribution of firms' productivities is contained in the cutoff productivity (for example, b_L easily translates into the output weighted average productivity). We explore this convenient feature further in the next section.

After collecting terms and inserting $\partial\epsilon/\partial\alpha = \epsilon^2$,

$$\frac{\partial \ln \pi\,(b, \cdot)}{\partial \alpha} = \frac{\epsilon}{\alpha} + \epsilon^2\left[2\ln\alpha - \ln(rb)\right].$$

Hence, profits are increasing in α if

$$\ln\frac{\alpha^2}{rb} > -\frac{1}{\alpha\epsilon},$$

or, using $-\alpha\epsilon = -\alpha/(\alpha-1)$,

$$\frac{\alpha^2}{rb} > e^{-\frac{1-\alpha}{\alpha}}.$$

Increasing α raises profits if $b < \bar{b}$, and lowers profits if $b > \bar{b}$, where

$$\bar{b} = \frac{\alpha^2 e^{\frac{1-\alpha}{\alpha}}}{r}.$$

For more productive firms (with $b < \bar{b}$), profits increase in α, since for them the positive effect of a high final good demand outweighs the negative effect due to a low mark-up. For less productive firms (those with input coefficient $b > \bar{b}$), the increase in demand is not sufficiently strong to outweigh the profit decreasing effect of a lower mark up. Since α captures opposing effects, comparative statics with respect to α are not unambiguous.

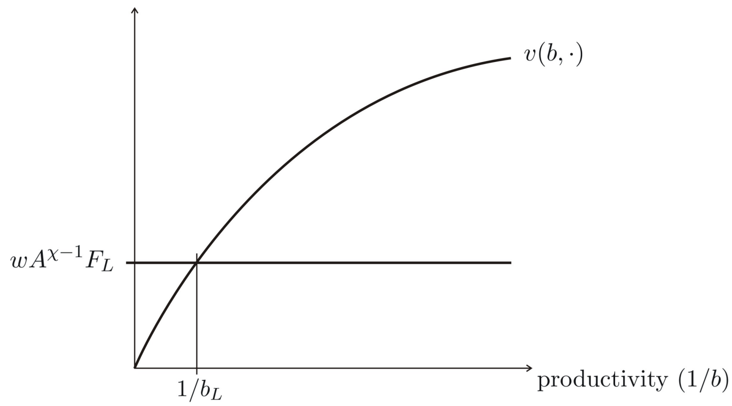

Figure 3: The cutoff productivity in autarky.

A law of motion for A. A binding cutoff ($b_L < b_0$) implies that researchers can only sell sufficiently productive blueprints to profit-seeking manufacturers. Given a continuum of newly discovered blueprints at any point in time, we rely on a law of large numbers and conclude that the fraction of profitable blueprints is $G(b_L)$. Hence, the evolution of A is governed by

$$\dot{A} = G(b_L)\dot{B} \text{ if } \dot{b}_L = 0. \tag{24}$$

Since only a fraction $G(b_L) < 1$ of newly discovered blueprint will actually go into production (and increase the specialization in the production of aggregate output), an increase in the minimum productivity requirement c.p. depresses the dynamic gains from horizontal specialization.

Labor allocation in R&E. In view of (24), let us clarify the allocation of labor between R&D and market entry. By construction, the ratio of labor in R&D to labor in entry is fixed for a given cutoff. From (3), (11), and (24),

$$\frac{L_B}{L_E} = \frac{F_B}{G(b_L)F_L}. \tag{25}$$

Every newly invented intermediate good requires F_L (times $A^{\chi-1}$) workers to realize its market entry and, on average, it takes $F_B/G(b_L)$ (times $A^{\chi-1}$) workers to discover a producible blueprint. Labor market clearing requires that the labor shares in entry, R&D, and production sum up to unity. Using this relation to replace L_B, and solving for the share of labor in the

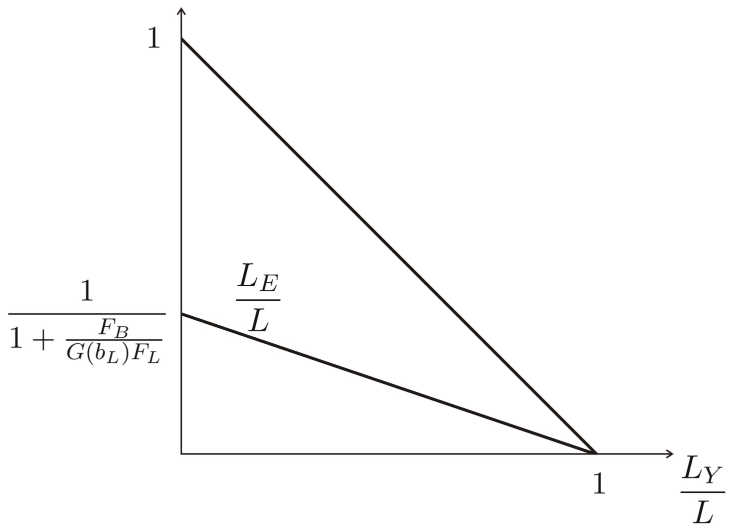

Figure 4: Labor shares in R&D and entry against the labor share in production for a given cutoff.

conduct of entry yields

$$\frac{L_E}{L} = \frac{1}{1+\frac{F_B}{G(b_L)F_L}}\left(1-\frac{L_Y}{L}\right). \tag{26}$$

Figure 4 shows the labor shares in R&E as a function of the labor share in the production of output for a given cutoff productivity $1/b_L$. The upper line depicts the labor market clearing condition as a function of the share of labor in production,

$$\frac{L_A}{L} = 1 - \frac{L_Y}{L}.$$

The lower line corresponds to the allocation of labor between entry and R&D, i.e. to equation (26). Of course, the horizontal distance between the two lines is the share of labor in R&D, L_B/L, since the labor market clearing line has slope -1. Suppose that the share of labor in production is not affected by the productivity distribution of intermediate firms (which we shall prove later on in Corollary 8). Then, for a given cutoff, $L_E/L(L_Y/L)$ simply centers around $L_Y/L = 1$ as F_L changes. We will come back to property after we will have characterized the equilibrium cutoff.

Free entry into R&D. In an equilibrium with free entry into R&D, the expected operating value net of market entry costs must at most outweigh the innovation cost. If $\dot{A} > 0$, we thus have

$$\int_0^{b_L} \left[v(b,\cdot) - wA^{\chi-1}F_L\right] dG(b) = wA^{\chi-1}F_B. \tag{27}$$

If the expected net return to R&D (the left-hand side), i.e. the market value of a capital good net of the entry cost (the term in squared brackets on the left-hand side), exceeds the R&D cost (the right-hand side), more researchers enter and discover a higher number of blueprints, thereby driving down the value of innovations. Similarly, if the expected net returns to R&D are not sufficient to cover the R&D cost, researchers leave and become production workers, thereby reducing the number of innovations and increasing the market value of innovations. Hence, the expected return to R&D must equal the total innovation costs; from (27),

$$\int_0^{b_L} v\left(b,\cdot\right) dG\left(b\right) = wA^{\chi-1}\left[F_B + G\left(b_L\right)F_L\right]. \tag{28}$$

Finally, we define an equilibrium.

Definition 1 (Equilibrium)**.** *An equilibrium is a path of quantities* $c, L_A, L_E, L_Y, Y, K, A, B, \{x(j), k(j)\}_{j\in[0,A]}$, *prices* $r, w, \{p(j), \pi(j), v(j)\}_{j\in[0,A]}$, *and the cutoff productivity* b_L *that satisfies technologies (1), (2), (3), (11), and (24), the entry conditions (23) and (27), the optimality conditions (12), (13), (14), and (15), the resource constraints (7) and (10), as well as the definitions of* π, v, *and* K.[27]

2.5 Aggregation for a given cutoff

We derive the equilibrium outcome in aggregate terms in two steps. First, we aggregate over all productivity type firms for a given level of the cutoff productivity. In a second step, we solve for the cutoff and characterize the equilibrium.

Suppose for the time being that the cutoff productivity $1/b_L$ is initially given and constant. Since all entrants are required to pay the entry costs, the productivity distribution in the product market, denoted by $\mu(b; b_L)$, is the productivity distribution of blueprints, $G(b)$, conditional

[27] As usual in general equilibrium theory, the households' budget constraint is another, but dependent, equation in the same variables.

on entry:[28]

$$\mu(b;b_L) \equiv \frac{G(b)}{G(b_L)} = \left(\frac{b}{b_L}\right)^{\theta}, \ b \in [0, b_L]. \tag{29}$$

Using $\mu(b)$, it is an easy task to aggregate over all active firm types.[29] Intuitively speaking, the probability density function $\mu'(b)$ gives the mass of firms for each level of productivity, relative to the total mass of active firms, A. The "number" of firms with the same level of productivity hence equals $A\mu'(b)$ for each productivity level b. Taking into account that only firms with productivities above the cutoff productivity incur the entry cost, integrating over all active productivity levels $b \le b_L$ then gives the aggregate intermediate outcome. To begin with, consider the capital stock in (9). Instead of aggregating over the raw capital inputs $k(j) = b(j)\,x(j)$ of all firms $j \in [0, A]$, we equivalently aggregate over all active productivity types $b \in [0, b_L]$, taking into account that there is a mass $A\mu'(b)$ of firms per level of productivity:

$$K = \int_0^A b(j)x(j)dj = \int_0^{b_L} bx(b,\cdot)\,A\mu'(b)\,db.$$

Now, using the conventional notation $d\mu(b) = \mu'(b)\,db$ and the equilibrium quantities from (16),

$$K = A\int_0^{b_L} b\alpha^{2\epsilon}(rb)^{-\epsilon}L_Y d\mu(b).$$

[28] In a setup with a random positive death rate for active firms and a large pool of potential entrants, Melitz (2003) shows that the long-run equilibrium distribution of active firms is $\tilde{G}(b)/\tilde{G}(b_L)$ if the universe of productivities is described by a more general class of probability distributions $\tilde{G}(b)$. The random death of firms of all productivity types is needed for the distribution of active productivity types to converge back to $\tilde{G}(b)/\tilde{G}(b_L)$ after a shock to b_L. In our environment, where A grows at a positive rate, the transition between two distributions of active firms' productivity types with different cutoffs is naturally achieved as the share of those productivities that are no longer introduced goes to zero in finite time. This is equivalent to randomized firm death, which steadily brings the productivity distribution of active firms back to the productivity distribution of newcomers whenever this distribution remains constant over time. To simplify the exposition, we drop the dependency of the active firms' productivity distribution's support on the cutoff whenever doing so does not lead to confusion.

[29] When aggregating over all firm types, we choose to express the outcome of the aggregation by equilibrium quantities of the cutoff productivity type firm. Following Melitz (2003), BRN, and GS we could alternatively apply an output-weighted average productivity type firm. Our choice, which of course is as good as any other productivity type, is motivated by the fact that the aggregate outcome in terms of the cutoff productivity makes the basic mechanism of the model visible very well.

After inserting

$$d\mu(b) = \frac{\theta b^{\theta-1}}{b_L^\theta} db \tag{30}$$

from (29) and integrating, the capital stock equals

$$K = \frac{A\alpha^{2\epsilon} r^{-\epsilon} L_Y \theta}{b_L^\theta} \int_0^{b_L} b^{\theta-\epsilon} db = \frac{A\alpha^{2\epsilon} r^{-\epsilon} L_Y \theta}{b_L^\theta} \left[\frac{b^{\theta-\epsilon+1}}{\theta - \epsilon + 1} \right]_0^{b_L} = A\alpha^{2\epsilon} r^{-\epsilon} L_Y \phi b_L^{1-\epsilon} \tag{31}$$

where $\phi \equiv \theta/(\theta - \epsilon + 1)$ (> 1).[30] To ease the exposition, use (16) again:

$$K = \phi A b_L x(b_L, \cdot). \tag{32}$$

The average productivity. The average output weighted productivity $\bar{b}$ is defined by

$$K = \bar{b} \int_0^A x(j) dj = \bar{b} A \int_0^{b_L} x(b, \cdot) d\mu(b).$$

Applying (30), inserting $x(b, \cdot)$ from (19), and integrating we have

$$K = \frac{\bar{b} A\alpha^{2\epsilon} \theta r^{-\epsilon}}{b_L^\theta} \int_0^{b_L} b^{\theta-1-\epsilon} db = \frac{\bar{b} \theta A\alpha^{2\epsilon} (r b_L)^{-\epsilon} L_Y}{\theta - \epsilon}.$$

Accordingly, using (19) and (32),

$$K = \frac{\theta A \bar{b} x(b_L, \cdot)}{\theta - \epsilon} = \frac{\theta A b_L x(b_L, \cdot)}{\theta - \epsilon + 1}, \tag{33}$$

so that

$$\bar{b} = \frac{b_L}{1 + \frac{1}{\theta - \epsilon}}. \tag{34}$$

For a given amount of accumulated savings, the output of intermediate firms is obviously larger, the more efficiently resources are transformed into intermediate goods, i.e. the smaller $\bar{b}$. Of course, with a Pareto distribution, the output-weighted average input coefficient increases with the input coefficient of the least productive firm.

Comparing the output-weighted average, $\bar{b} = (\theta - \epsilon)/(\theta + 1 - \epsilon) b_L$, to the unweighted average which corresponds to symmetric varieties,

$$\int_0^{b_L} b d\mu = \frac{\theta \int_0^{b_L} b^\theta db}{b_L^\theta} = \frac{b_L}{1 + \frac{1}{\theta}},$$

[30] $\phi = \frac{\theta}{\theta - (\epsilon - 1)} > 1$ since $\theta > \epsilon - 1$ by (PA1) and $\epsilon > 1$.

confirms the intuition that the difference in firms' output is more pronounced, the lower the degree of substitutability between intermediate goods ($\epsilon \uparrow$). That is, competition in the product market (measured by the degree of substitutability between intermediate goods) works against the variance reducing effect of fixed cost.[31]

Aggregate profits and firm values. Turning to firm's market values, we first derive the aggregate intermediate producers' profits. Using (15) and (31),

$$\int_0^A \pi(j)\,dj = (1-\alpha)\int_0^A p(j)\,x(j)\,dj = \frac{(1-\alpha)\,r}{\alpha}\int_0^A b(j)\,x(j)\,dj = (1-\alpha)\,\alpha^{2\epsilon-1}\phi A\,(rb_L)^{1-\epsilon}\,L_Y.$$

From (17), we have

$$\int_0^A \pi(j)dj = (1-\alpha)A\phi R(b_L,\cdot) = A\phi\pi(b_L,\cdot). \tag{35}$$

The average profit is thus ϕ times the profit of firms operating with the cutoff productivity, $\phi\pi(b_L) = \int_0^A \pi(j)dj/A$. Using (6), the same is true for the cutoff productivity type firm value. From $\pi(j) = v(j)(r-\hat{v})$ and (35), we find

$$\int_0^A v(j)dj = A\phi v(b_L,\cdot). \tag{36}$$

The difference in the market value of firms with the cutoff productivity and the average productivity is larger, the larger ϕ. ϕ accounts for the characteristics of the underlying distribution of productivities (as summarized by θ and b_0) and includes α as an indicator of the value of productivity.[32] Consistent with the previous observation on relative profits, the value of average productivity type firms is low relative to the value of firms operating with the cutoff productivity if α is small (ϕ is larger, the larger α). A large α implies a high level of all firms' values[33], and more unevenly distributed profits. Put differently, the dispersion in the values of firms with different productivity levels depends positively on α ($\alpha \to 0$ implies $\phi \to 1$ and

[31] Note that this conclusion is again ambiguous due to the fact that α also measures the share of capital income and the gains from specialization.

[32] Of course, a higher average productivity implies a higher distance of the average to the cutoff, so ϕ increases with the breadth and dispersion of the underlying distribution, b_0 and θ.

[33] This is because profits are higher the larger α, see the paragraph below equation (17).

$v(b,\cdot) \to v(b_L,\cdot))$.[34]

Next, aggregating over the intermediate firm's outputs in (1) using the equilibrium quantities from (19) and $d\mu\left(b\right)$ from (30), the production function for aggregate output can be rewritten as

$$Y = L_Y^{1-\alpha}\int_0^A x(j)^\alpha dj = L_Y^{1-\alpha}A\alpha^{2\epsilon\alpha}r^{1-\epsilon}L_Y^\alpha\int_0^{b_L} b^{1-\epsilon}d\mu(b) = L_Y^{1-\alpha}\phi A[\alpha^{2\epsilon}(rb_L)^{-\epsilon}L_Y]^\alpha,$$

or, using (16) again,

$$Y = AL_Y^{1-\alpha}\phi x(b_L,\cdot)^\alpha. \tag{37}$$

Replacing $x(b_L)$ using (32) we find

$$Y = (\phi AL_Y)^{1-\alpha}\left(\frac{K}{b_L}\right)^\alpha = \phi(AL_Y)^{1-\alpha}\left(\frac{K}{\phi b_L}\right)^\alpha \tag{38}$$

Equation (38) demonstrates quite clearly the close analogy to a representative firm model, where $Y = (AL_Y)^{1-\alpha}K^\alpha$. In the present environment with heterogeneous firms, the (endogenously increasing) degree of specialization is complemented by the (static) average input coefficient that describes how efficient capital goods can be manufactured to produce output. A degenerate one point distribution at $b = b_L$ ($\theta \to \infty$) implies $\phi \to 1$.[35]

The output-capital ratio. We already know from the household's optimal consumption decision, that the savings behavior is not affected by the productivity distribution of intermediate firms. Hence, the output-capital ratio should be independent of the distribution of firm productivities. From (38),

$$\frac{Y}{K} = \frac{(\phi AL_Y)^{1-\alpha}K^{\alpha-1}}{b_L^\alpha},$$

and using (31),

$$K^{\alpha-1} = A^{\alpha-1}\alpha^{2\epsilon(\alpha-1)}r^{-\epsilon(\alpha-1)}L_Y^{\alpha-1}\phi^{\alpha-1}b_L^{(1-\epsilon)(\alpha-1)}.$$

[34]The latter observation is easily verified by looking at relative firm values. For $b < b_L$,

$$\frac{d\left[\frac{v(b,\cdot)}{v(b_L,\cdot)}\right]}{d\alpha} = \frac{d\left[\frac{\pi(b,\cdot)}{\pi(b_L,\cdot)}\right]}{d\alpha} = \frac{d\left(\frac{b_L}{b}\right)^{\frac{\alpha}{1-\alpha}}}{d\alpha} = \left[\frac{1}{1-\alpha} + \frac{\alpha}{(1-\alpha)^2}\right]\left(\frac{b_L}{b}\right)^{\frac{\alpha}{1-\alpha}}\log\left(\frac{b_L}{b}\right) > 0.$$

[35]For $\theta \to \infty$, $\mu(b_L) = 0$ for all $b < b_L$ and $\mu(b_L) = 1$ for $b = b_L$. At the same time, $\lim_{\theta\to\infty}\frac{1}{1-\frac{\epsilon-1}{\theta}} = 1$.

By definition, $\epsilon(\alpha-1)=-1$ and $(1-\epsilon)(\alpha-1)=\alpha$, so that

$$\frac{Y}{K}=\frac{r}{\alpha^2}. \tag{39}$$

We thus note:

Result 2. *The output-capital ratio depends positively on the interest rate. Unless entry costs have an impact on the interest rate, the output-capital ratio is independent of barriers to entry.*

The evolution of A for a given cutoff. Given b_L, a compact law of motion for A is readily obtained by combining (24), (3), and (11). From the first two equations,

$$\dot{A}=\frac{L_B A^{1-\chi}G(b_L)}{F_B}=\frac{(L_A-L_E)A^{1-\chi}G(b_L)}{F_B}.$$

Inserting L_E from (11) yields

$$\dot{A}=\frac{(L_A-\dot{A}A^{\chi-1}F_L)A^{1-\chi}G(b_L)}{F_B}.$$

Solving for $\dot{A}$, the R&E process is described by a standard Jones (1995) R&D technology:

$$\dot{A}=\frac{L_A A^{1-\chi}}{F^a(b_L)}, \quad F^a(b_L)\equiv F_L+\frac{F_B}{G(b_L)}. \tag{40}$$

Compared to Jones (1995), the innovation technology is augmented in two aspects. Firstly, without entry costs, all research attempts are successful. Here, the R&E productivity $(1/F^a(b_L))$ decreases endogenously with b_L because some innovations must be discarded. $F^a(b_L)=F_L+F_B/G(b_L)$ captures the effect of entry costs and the implied minimum productivity requirement on the R&D productivity in terms of output quantities. Given A, discovering and launching production for a new intermediate good is obviously less labor intensive if the minimum productivity requirement is low, or easy to meet (because θ is high so that $G(b_L)$ is high), and if few workers are necessary to conduct market entry (i.e. if F_L is low). Note that without entry cost (more precisely, with F_L violating (PA1)), there is no need to dispense with low productivity types. Here, in contrast, it takes $1/G(b_L)$ times more resources on average to discover a usable blueprint. Secondly, entry is modeled in such a way that it takes workers away from R&D. This further increases the labor requirement necessary for usable blueprints.

Free entry in R&D for a given cutoff. Diving the free entry into innovation condition in (28) by $G(b_L)$ gives

$$\int_0^{b_L} v(b,\cdot)\frac{dG(b)}{G(b_L)} = wA^{\chi-1}F^a(b_L).$$

After recognizing

$$\frac{dG(b)}{G(b_L)} = \frac{G'(b)\,db}{G(b_L)} = \mu'(b)\,db = d\mu(b),$$

and substituting

$$v(b,\cdot) = \frac{\pi(b,\cdot)}{\left[r - \frac{\dot{v}(b,\cdot)}{v(b,\cdot)}\right]} \tag{41}$$

from (6) in terms of input coefficients (see (20)) we get

$$\int_0^{b_L}\left[\frac{\pi(b,\cdot)}{r - \frac{\dot{v}(b,\cdot)}{v(b,\cdot)}}\right] d\mu(b) = wA^{\chi-1}F^a(b_L).$$

As shown before, $\hat{v}(b,\cdot) = \hat{v}$, hence $r - \hat{v}$ is independent of b and can be pulled out of the integral. Moreover, since

$$\int_0^{b_L} \pi(b,\cdot)d\mu(b) = \frac{\int_0^A \pi(j)dj}{A},$$

we can replace the remaining integral term, the average profits, with the expression implied by (35), i.e. $\phi\pi(b_L,\cdot)$. Finally, using (41), the free entry into R&D condition becomes

$$\phi v(b_L,\cdot) = wA^{\chi-1}F^a(b_L). \tag{42}$$

The right-hand side equals the average development costs of an actually producible durable good: a newly discovered blueprint requires F_L times $A^{\chi-1}$ workers to conduct its market entry and it takes $F_B/G(b_L)$ times $A^{\chi-1}$ workers on average to discover a producible blueprint in the first place (researchers on average must "draw" $1/G(b_L)$ times to find a sufficiently productive type). In endogenenous growth models with free entry into innovation and costless entry into the product market, the R&D costs of every undertaken research project must equal its costs in an equilibrium with positive growth. In the present environment, however, researchers face uncertainty about the productivity and thus the market value of their innovations. In particular, blueprints with a productivity below the cut-off will not earn their R&D costs. Hence, in equilibrium, sucessful innovations must earn excess rents. Inserting the definition of F^a from (40) in (42) and solving for the R&D costs of a single discovery shows this most clearly:

$$wA^{\chi-1}F_B = G(b_L)\left[\phi v\left(b_L,\cdot\right) - wA^{\chi-1}F_L\right] < \phi v\left(b_L,\cdot\right) - wA^{\chi-1}F_L.$$

Given that the cutoff is binding, i.e. $G(b_L) < 1$, the average net value of entry (the right-hand side of the inequality) exceeds the actual R&D costs of a single innovation to ensures that research investments break even across all undertaken projects. More generally, if there is a positive probability that research projects fail, the ex post return on sucessful projects must exceed one, to ensure free entry ex ante. Since this feature is the main difference between the heterogeneous firms and entry costs models and the canonical growth models, we explicitly state it in

Result 3 (Excess rents for innovators)**.** *The average net value of entry exceeds the innovation cost.*

Return on investment in R&D. To avoid confusion, we explicitly state that ex ante zero profits free entry in R&D imply that the ratio of the average firm value to the average R&D costs for a producible blueprint is independent from the entry costs. From (42),

$$\frac{\phi v(b_L,\cdot)}{wA^{\chi-1}F^a(b_L)} = 1.$$

Whenever the R&D costs should increase as a consequence of increasing entry costs, the average returns to successful R&D would increase by the same factor.

Recap. Let us recapitulate briefly. Melitz (2003) showed that dealing with firm heterogeneity is easy when consumers have love of variety preferences à la Dixit-Stiglitz because these preferences still allow us to work with a single, representative firm. The same is true if we follow Ethier (1982) and use a Dixit-Stiglitz aggregator in the production of output. To reveal the basic mechanics of the model, we choose to express the aggregate firm outcome in terms of the cutoff productivity type firms. Given the cutoff, R&E is conducted with a standard Jones (1995) R&D technology. In fact, the closed economy model with heterogeneous firms and entry costs boils down to the Jones (1995) model. The two additional ingredients, costly entry and firm heterogeneity, are included as follows. In Jones' model, the R&D technology is

$$\dot{A} = \frac{A^{1-\chi}L_A}{a},$$

where a is an exogenously given productivity parameter. Including entry costs, we can interpret the productivity as being endogenous. In our formulation, it incorporates the labor requirement necessary for market entry and to find a sufficiently productive blueprint (a in Jones' model can take any value so we set $a \equiv F^a(b_L)$).[36]

The aggregate equilibrium outcome with firm heterogeneity can conveniently be expressed as the outcome with a representative firm.Market entry costs introduce a minimum productivity requirement that increases the average productivity of firms. The net effect of entry costs on the level of TFP, however, is ambiguous, since an increase in productivity in production comes at the cost of an increase in the average labor requirement necessary to invent a new variety. If the share of labor in R&D remains fixed, this increase translates into a lower rate at which new intermediate goods are introduced to the output sector. Note, however, that we cannot assert that the R&D costs actually increase until we know more about the effects of the entry costs on the wage rate and the level of A which governs the spillover effects.

Returning to the derivation of an equilibrium, it remains for us to solve for the lowest productivity level that allows firms to earn the entry costs (the cutoff productivity).

2.6 The equilibrium cutoff productivity

Our choice of expressing the aggregate intermediate firm outcome in terms of the cutoff productivity type firms shows clearly that solving for the cutoff productivity requires only the free entry into R&D condition and the condition for profitable market entry. To see this, recall that the free entry condition requires that the average productivity type firms' values net of entry costs are equal their R&D costs. The value of firms with the average productivity in turn is closely linked to the cutoff productivity firms' values, see (36) for a given A. By definition, the cutoff productivity type firms' net/market value in turn is zero, i.e. their operating value equals the entry costs. The equilibrium cutoff productivity must therefore imply an average firm value that exactly meets the average R&D costs of finding a usable blueprint. Combining (23) and

[36] L_A in our formulation also includes entry workers (which do not exist in Jones' model).

(42) yields

$$F^a(b_L^*) = \phi F_L, \tag{43}$$

or from (40),

$$G(b_L^*) = \frac{F_B}{F_L(\phi - 1)}. \tag{44}$$

The wage rate and the scaling factor $A^{\chi-1}$ drop out because of identical technologies in R&D and market entry. Hence, free entry into R&D requires the average net value of a profitably usable blueprint $F_L(\phi-1)wA^{\chi-1} = (\phi F_L - F_L)\,wA^{\chi-1} = \phi v_L wA^{\chi-1} - F_L wA^{\chi-1}$ times the fraction of usable blueprints $G(b_L^*)$, to equal the discovery costs $F_B wA^{\chi-1}$. Put differently, researchers may expect a usable blueprint after $1/G(b_L^*) = F_L(\phi-1)/F_B$ draws on average. If successful, the return on the research investment equals $F_L(\phi-1)/F_B$, so that, in expectation, researchers exactly break even on average. Since blueprints with productivities below the cutoff have zero value, the share of usable discoveries is equal to the inverse of the return on investment in R&D for any usable blueprint.

From the definition of $G(b)$ in (4), $b_L^* = [G(b_L)^*]^{\frac{1}{\theta}}\, b_0$. Inserting (44) yields the equilibrium cutoff input coefficient:

$$b_L^* = \left[\frac{F_B}{F_L(\phi-1)}\right]^{\frac{1}{\theta}} b_0 \quad (> 0). \tag{45}$$

The separation of firms into profitable producers and firms that exit comes solely from the entry costs. As mentioned earlier, $b_L^* \to \infty$ as $F_L \to 0$ (i.e. the cutoff is not binding as b_0 is finite, $b_L^* = b_0$). The minimum productivity requirement, $1/b_L^*$, is obviously higher, the higher F_L.[37] Interestingly, the efficiency with which researchers operate to find new blueprints ($1/F_B$) has a negative impact on the highest admissible input coefficient: the cutoff input coefficient is smaller, the more efficient the development of new blueprints occurs (i.e. the smaller F_B).

[37] The closed economy model in this chapter merely serves as a starting point for the analysis of marginal changes in the openness of foreign markets (as indicated by the foreign market entry costs). While the above comparative static is helpful to understand the model's mechanics, we do not want to take the productivity-increasing effect of local market entry costs too serious. This is because sunk entry costs (like, e.g., costly regulation) in general deter the creation of new firms, a feature broadly supported by the data, see Alesina et al. (2005), Nicoletti and Scarpetta (2003), and Klapper et al. (forthcoming). Loosely speaking, the average firm in Greece, where entry costs are about US$ 6900, is hardly believed to be more productive than the average firm in Canada, where entry cost are much lower (US$ 280 according to Buettner, 2006).

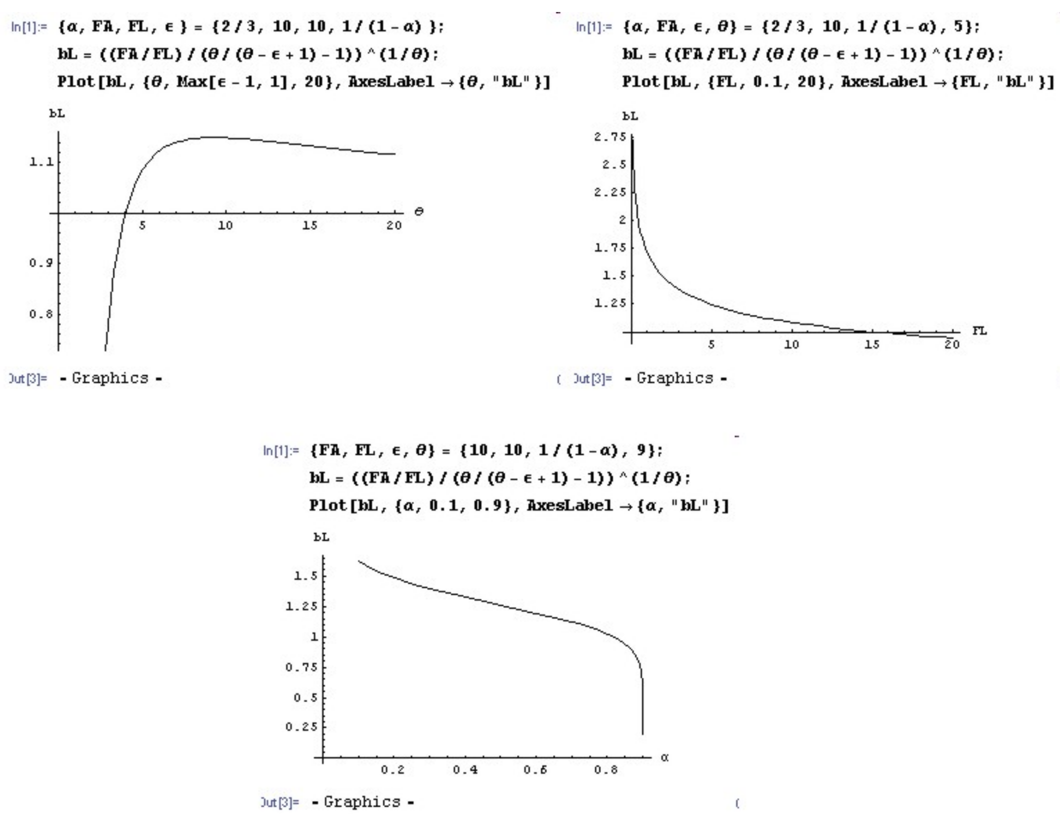

Figure 5: b_L^* as a function of θ (upper left panel), F_L (upper right panel), and α (lower panel).

In other words, an R&D sector with a low productivity allows intermediate firms to be less efficient in production (horizontal and vertical productivity are complements). This is intuitive because a less productive R&D sector implies less pronounced horizontal competition from new entrants (and higher profits for incumbent firms). Figure 5 depicts the cutoff input coefficient as a function of θ (upper left panel), F_L (upper right panel), and α (lower panel). We summarize these findings in

Result 4 (Minimum productivity requirement)**.** *Entry barriers introduce a minimum productivity requirement for intermediate goods firms. This requirement is higher when the capital share in the production of output is large (when α is large) and when researchers are productive in the discovery of blueprints (when F_B is small).*[38]

Note that since the factor prices drop out in the determination of the cutoff, see (43), the

[38]Given the different meanings of α, the result with respect to the capital share is again ambiguous.

production structure is not essential for the determination of the cutoff when the production functions in R&D and entry are identical.

2.7 Properties of the autarky equilibrium

We are now in a position to characterize the equilibrium labor allocation and the evolution of horizontal specialization.

Irrelevance of entry costs for the allocation of labor in R&E. Starting with the labor allocation, we find that the relative inputs of labor in R&D and entry are not affected by barriers to entry as measured by F_L. This at first glance astonishing feature is concealed in existing models, where there is no explicit distinction between entry workers and researchers. BRN and GS take a short cut by assuming that the discovery of new intermediates takes a certain amount of knowledge and that it takes an additional amount of knowledge to enter a market subsequently. As a consequence, they directly employ an R&E-knowledge production function like (40). Clearly, we we do not alter the modeling substantially given that we maintain the mechanical link between entry workers and researchers implied by identical production functions. Exploring the relation between entry workers and researchers, however, reveals how restrictive this assumption actually is: in equilibrium, the allocation of labor between market entry and R&D is not affected by the entry barriers.

To see this, plug

$$G(b_L^*)F_L = \frac{F_B}{\phi - 1}$$

from (44) into (25):

$$\frac{L_B}{L_E} = \phi - 1. \tag{46}$$

From $L_B + L_E \equiv L_A$, $L_B = (\phi - 1)(L_A - L_B)$ such that

$$\frac{L_B}{L_A} = \frac{\phi - 1}{\phi}, \tag{47}$$

$$\frac{L_E}{L_A} = \frac{1}{\phi}. \tag{48}$$

Hence, the existence of a binding cutoff is sufficient to fix the relative labor shares in R&D and entry, irrespective of the level of the barriers to entry. At second glance of course, the increase in the labor requirement per usable blueprint is only one side of the coin. Turning the minimum productivity requirement up side down, barriers to entry reduce the share of newly invented, usable blueprints. In equilibrium, a change in the barriers to entry induces a change in the share of usable blueprints that exactly offsets the change in the labor requirement for developing a usable blueprint. Formally, from (44), $G(b_L^*)F_L$ is fixed independent of F_L.

More explicitly, lowering F_L has two opposing effects. On the one hand, a reduction in the barriers to entry frees labor from the conduct of entry for a given number of innovations (i.e. L_A/L_E rotates counter-clockwise around $L_Y/L = 1$ in Figure 4). On the other hand, the number of newly invented intermediate goods increases for a given number of researchers because the reduction in entry costs relaxes the minimum productivity requirement (and thereby reduces the average labor requirement necessary to discover a usable blueprint). Hence, there are more usable innovations for which the free entry workers conduct market entry. If there is a change in the share of researchers as a fraction of the labor force, it is accompanied by an equally sized change in the share of entry workers.

Result 5 (Labor allocation between R&D and entry)**.** *In equilibrium, the relative use of labor in R&E, L_E/L_B, is independent of the level of the entry costs.*

Law of motion for A**.** Using $F^a(b_L^*)$ from (43) in (40) gives the equilibrium evolution of A:

$$\dot{A} = \frac{A^{1-\chi}L_A}{\phi F_L}. \tag{49}$$

Barriers to entry decrease the rate at which new blueprints are introduced to the production of output, and their impact is stronger the easier it is for the producers of output to replace inputs of less efficient firms by (cheaper) inputs of more efficient firms (as ϕ is increasing in ϵ).

The underlying R&D productivity ($1/F_B$) drops out because the productivity of blueprints used in active firms is conditional on exceeding the cutoff. On average, the discovery of these blueprints requires $F_B/G(b_L^*)$ (times $A^{\chi-1}$) workers. Since the probability of drawing a productivity of at least $1/b_L^*$ in equilibrium always takes the same effort ($G(b_L^*) = F_B/[F_L(\phi-1)]$ is linear in F_B), conditioning on $b \leq b_L^*$ removes F_B from the law of motion for A.

Result 6 (Irrelevance of F_B for $\dot{A}$)**.** *In equilibrium, the law of motion for A is pinned down by the entry costs irrespective of the productivity in R&D.*

Equation (40) indicates that the reduced form of our model yields the Jones (1995) model. We will verify this conjecture explicitly in the following section. For now, note that we are free to choose the units of measurement in the production of output, so that the production function of output in (1) can equivalently be stated as

$$\tilde{Y} = \delta L_Y^{1-\alpha} \int_0^A x(j)^\alpha dj, \ \delta > 0, \ \dot{\delta} = 0. \tag{50}$$

Inserting the equilibrium cutoff from (45) in the reduced form production function of output in (38) gives

$$Y = \phi (A L_Y)^{1-\alpha} \left\{ \frac{K}{\phi \left[\frac{F_B}{F_L(\phi-1)} \right]^{\frac{1}{\theta}} b_0} \right\}^\alpha = \left[\frac{F_L(\phi-1)}{F_B} \right]^{\frac{\alpha}{\theta}} \frac{\phi^{1-\alpha}}{b_0^\alpha} (A L_Y)^{1-\alpha} K^\alpha.$$

Without loss of generality for a given F_L let

$$\delta \equiv \left[\frac{F_B}{F_L(\phi-1)} \right]^{\frac{\alpha}{\theta}} \frac{b_0^\alpha}{\phi^{1-\alpha}}$$

so that we get a standard Cobb-Douglas production function for output:

$$\tilde{Y} = (A L_Y)^{1-\alpha} K^\alpha. \tag{51}$$

By definition, more productive firms produce more output out of a given amount of input. Hence, δ increases as F_L decreases since the average productivity depends positively on the monotonically increasing minimum productivity requirement introduced by entry costs.

2.8 Balanced Growth Path

Given that the cutoff is determined by instantaneous optimality conditions, the dynamics of the model are identical to the dynamics of the Jones (1995) model (which are analyzed in Arnold, 2006). For the sake of completeness, we adapt the analysis in Arnold (2006) to the present environment where intermediate firms have heterogeneous input coefficients, but can be summarized by a representative firm. We then use the laws of motions of key variables to solve for the equilibrium allocation along a balanced growth path. Following Arnold (2006), define the stationary variables $\tilde{l} \equiv L/A^\chi$, $z \equiv Y/K$, $\gamma \equiv cL/K$, and $\nu \equiv (1-\alpha)Y/[\int_0^A v(j)dj] = (1-$

$\alpha)Y/[A\phi v(b_L^*)$. We use $\tilde{l}$ instead of $l \equiv L/\left[\phi F_L A^{\chi}\right] = \tilde{l}/\left(\phi F_L\right)$ to trace the R&E productivity in the law of motion for A. Of course, we can alternatively derive the balanced growth path without these definitions.[39] Deriving the entire dynamic system, however, increases our understanding of the model's mechanics and gives us an idea of the interdependencies off the balanced growth path.

Definition 2 (Steady state)**.** *A steady state is an equilibrium with l, z, γ and ν constant.*

Before we consider the dynamic system, we first report the steady state growth rates of all endogenous variables.

2.8.1 Steady state growth rates

By definition of a steady state, all variables grow at constant rates. Imposing a constant growth rate on consumption requires, via the Euler equation (12), that the interest rate is a constant as well. Accordingly, from (15), the prices of the intermediate goods are constant. From labor market clearing, labor in all sectors must grow at the population growth rate so as to ensure that the labor shares are constant. Profit maximization in the production of output then implies that the input quantities of intermediate goods also grow at rate n, see (14). As prices and the interest rate are constant, the same holds true for profits and firm values, see (18) and (22). Then, using the results of the aggregation in (32) and (51), the capital stock and final output grow at rate $n + n/\chi$. For the growth rate of blueprints to be a constant, B must grow at rate n/χ.[40] As b_L^* is a constant (see (45)), (24) demands that A is a constant fraction of B. Hence, A grows at the same rate as B. Finally, the profit maximizing labor demand in the production of output in (13) implies that the wage rate grows at rate n. To summarize, $\hat{A} = \hat{B} = \hat{w} = \hat{c} = n/\chi$, $\hat{K} = \hat{Y} = n + n/\chi$, $\hat{L}_Y = \hat{L}_E = \hat{L}_B = \hat{x} = \hat{\pi} = \hat{v} = n$, and $\hat{r} = \hat{p} = \hat{b}_L^* = 0$.

[39]As an example, consider ν. ν^* is simply derived from the long-run values of r and $\hat{v}^*$. $\nu = \frac{(1-\alpha)Y}{A\phi v(b_L,\cdot)} = \frac{\phi A \pi_L}{\alpha \phi A v_L}$ since $A\phi\pi_L = (1-\alpha)\alpha Y$. From (6) in a steady state, $r^* - \hat{v} = \pi(j)/v(j) = \pi\left(b_L,\cdot\right)/v\left(b_L,\cdot\right)$. Substituting for the dividend ratio delivers ν^*.

[40]In (40), $\dot{A} = \frac{A^{1-\chi}L_A}{\phi F_L}$, so that $\dot{\hat{A}} = 0$ requires $\hat{A} = \frac{n}{\chi}$. Hence, (24) and (44) yield $\hat{B} = \hat{A} = \frac{n}{\chi}$.

In a steady state, the growth rate of consumption per capita, $\hat{c}/c = n/\chi$, and the Euler equation in (12) pin down the interest rate irrespective of the barriers to entry:

$$r^* = \sigma\frac{n}{\chi} + n + \rho. \tag{52}$$

That is, the long-run interest rate is such that it removes the dissaving motives from population growth, subjective discounting, and growth (which is larger the smaller the elasticity of intertemporal substitution).[41]

Result 7 (Optimal consumption determines steady state interest rate)**.** *In a steady state, the interest rate, i.e. the factor price for intermediate goods firms, is independent of barriers to entry.*

Impact channels of entry costs. While this relation is generally well known for the Jones (1995) model, it is an important observation with respect to the impact of trade. Intuitively speaking, a reduction in the barriers to trade exerts an production-expanding effect (via additional entry and production for foreign markets by exporters) which is expected to bid up the factor price. In fact, in the Melitz (2003) model, the trade-induced increase in the real wage rate is the only channel though which trade openness drives the least productive firms out of the domestic market (by increasing the local market's minimum productivity requirement). In particular, the constant price elasticity of demand implied by the Dixit-Stiglitz index severely limits the possible impact of trade on factor price effects. Changing the number of competitors or their productivity leaves the elasticity of demand unaffected (see Melitz, 2003, p. 1715). In the present model, Result 7 implies that the steady state factor price is independent of the minimum productivity requirement implied by the entry costs. In looking for the impact of trade in a monopolistic competition model under CES production, this observation hints at looking for decreases in the prices for the other production factors, labor and knowledge.

In what follows, we deduce the laws of motions and the steady state values for the transformed variables. If one is less interested in this rather technical derivation, one can skip this

[41]From the point of view of the market for the investment good, the determinants of r equivalently reflect the scarcity of raw capital.

paragraph. The stationary values of key variables are identical to those in the Jones (1995) model since the minimum productivity requirement only enters through the productivity in R&E.

2.8.2 Laws of motions for key variables

To economize on notation, let $\eta(b_L^*) = \eta_L$, $\pi(b_L^*, \cdot) = \pi_L$ and $v(b_L^*, \cdot) = v_L$. Log-differentiating the definition of γ, we have

$$\dot{\gamma} = \gamma\left(\frac{\dot{c}}{c} + n - \frac{\dot{K}}{K}\right).$$

To substitute for $\dot{c}/c$ and $\dot{K}/K$, note from the output-capital ratio in (39) that

$$\alpha^2 z = r.$$

Dividing by K, the resource constraint in (10) implies $\dot{K}/K = z - \gamma$. Using these expressions together with the Euler equation yields the law of motion for γ:

$$\dot{\gamma} = \gamma\left[\left(\frac{\alpha^2}{\sigma} - 1\right)z + \gamma + \left(1 - \frac{1}{\sigma}\right)n - \frac{\rho}{\sigma}\right]. \tag{53}$$

The aggregate firm value is equal to $\int_0^A v(j)dj = \phi A v_L$ (see (36)). As $v_L(r - \hat{v}) = \pi_L$ from (22) and (20), the denominator of $\nu = (1-\alpha)Y/(\phi A v_L)$ is

$$\phi A v_L = \frac{\phi A \pi_L}{r - \hat{v}}. \tag{54}$$

Aggregate profits from (35), rewritten as

$$\phi A \pi(b_L) = \int_0^A \pi(j)dj = (1-\alpha)\alpha^{2\epsilon-1}\int_0^A [rb(j)]^{1-\epsilon} L_Y dj = (1-\alpha)\alpha L_Y^{1-\alpha}\int_0^A \alpha^{2\alpha\epsilon}[rb(j)]^{1-\epsilon} L_Y^{\alpha} dj,$$

can be expressed as a constant fraction of aggregate output. Recognizing $-\alpha\epsilon = 1-\epsilon$ and using (16),

$$A\phi\pi_L = (1-\alpha)\alpha\int_0^A x(j)^{\alpha} dj = \alpha(1-\alpha)Y. \tag{55}$$

Taken together, (54) and (55) imply $r - \hat{v} = \alpha\nu$. Replacing $r = \alpha^2 z$ gives the law of motion for the intermediate firms' values:

$$\hat{v} = \alpha(\alpha z - \nu). \tag{56}$$

The equilibrium law of motion for A from (49) can be rewritten using the labor market clearing condition (7), the equilibrium demand from the aggregate output sector (14), and the free entry condition (42), which equals the free entry into the product market condition in equilibrium:

$$\frac{\dot{A}}{A}=\frac{A^{-\chi}L_A}{\phi F_L}=\frac{L-L_Y}{A^{\chi}\phi F_L}=\frac{\tilde{l}}{\phi F_L}-\frac{(1-\alpha)YF_L}{A^{\chi}\phi F_L v_L A^{1-\chi}}=\frac{\tilde{l}}{\phi F_L}-\nu, \tag{57}$$

or simply $\dot{A}/A=l-\nu$.

Now, combining (32) and (37), we have $Y/K=\phi\varphi^{-\alpha}(AL_Y)^{1-\alpha}K^{\alpha-1}$, or

$$z=\frac{\phi}{\varphi^{\alpha}}\left(\frac{AL_Y}{K}\right)^{1-\alpha}. \tag{58}$$

From (13) and $L_Y=(1-\alpha)Y/w$, (58) becomes

$$z=\frac{\phi}{\varphi^{\alpha}}\left[\frac{(1-\alpha)AY}{wK}\right]^{1-\alpha}.$$

Solving for $z=Y/K$ yields

$$z=\frac{\phi^{\frac{1}{\alpha}}}{\varphi}\left[\frac{(1-\alpha)A}{w}\right]^{\frac{1-\alpha}{\alpha}}.$$

We can replace the wage rate in the above expression from the free entry condition (which in equilibrium coincides with the entry into the product market condition), $w=v_L A^{1-\chi}/F_L$, and get

$$z=\frac{\phi^{\frac{1}{\alpha}}}{\varphi}\left[\frac{(1-\alpha)A^{\chi}F_L}{v_L}\right]^{\frac{1-\alpha}{\alpha}}. \tag{59}$$

Log-differentiating yields

$$\dot{z}=z\frac{1-\alpha}{\alpha}\left(\chi\frac{\dot{A}}{A}-\frac{\dot{v}}{v}\right),$$

or, replacing $\dot{A}/A$ from (57) and $\dot{v}/v$ from (56),

$$\dot{z}=z\left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\chi}{\phi F_L}\tilde{l}+(\alpha-\chi)\nu-\alpha^2 z\right]. \tag{60}$$

Turning to $\tilde{l}$, $\dot{\tilde{l}}=\tilde{l}(g-\chi\dot{A}/A)$, and inserting $\dot{A}/A$ from (49), one obtains

$$\dot{\tilde{l}}=\tilde{l}\left[n-\chi\left(\frac{\tilde{l}}{\phi F_L}-\nu\right)\right]. \tag{61}$$

Finally, a differential equation for ν is obtained as follows. Dividing (10) by K, we get the law of motion for the capital stock,

$$\frac{\dot{K}}{K}=(z-\gamma). \tag{62}$$

Using the definition of z, $\dot{Y}/Y = \dot{z}/z + \dot{K}/K$, and after replacing $\dot{K}/K$ from (62) and $\dot{z}/z$ from (60),

$$\frac{\dot{Y}}{Y} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\chi}{\phi F_L}\tilde{l} + (\alpha-\chi)\nu - \alpha^2 z\right] + z - \gamma. \tag{63}$$

Now plugging this expression together with the laws of motion for $v(j)$ and A in (56) and (57) in the log-differentiated definition of ν, $\dot{\nu} = \nu(\hat{Y} - \hat{A} - \hat{v})$,

$$\frac{\dot{\nu}}{\nu} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\chi}{\phi F_L}\tilde{l} + (\alpha-\chi)\nu - \alpha^2 z\right] + z - \gamma - \frac{\tilde{l}}{\phi F_L} + \nu - \alpha(\alpha z - \nu).$$

After collecting terms, the law of motion for ν equals

$$\dot{\nu} = \nu\left[\left(\frac{1-\alpha}{\alpha}\chi - 1\right)\frac{\tilde{l}}{\phi F_L} + \left(2 - \frac{1-\alpha}{\alpha}\chi\right)\nu + (1-\alpha)z - \gamma\right]. \tag{64}$$

2.8.3 Steady state

In a steady state, where $\dot{\gamma} = \dot{z} = \dot{\tilde{l}} = \dot{\nu} = 0$, (53), (60), (61), and (64) imply

$$\begin{aligned}
\left[\left(\frac{\alpha^2}{\sigma} - 1\right)z^* + \gamma^*\right] + \left(1 - \frac{1}{\sigma}\right)n &= \frac{\rho}{\sigma} && (65)\\
\frac{\chi}{\phi F_L}\tilde{l}^* + (\alpha-\chi)\nu^* &= \alpha^2 z^* && (66)\\
\frac{\chi}{\phi F_L}\tilde{l}^* &= n + \chi\nu^* && (67)\\
\left(\frac{1-\alpha}{\alpha} - \frac{1}{\chi}\right)\frac{\chi}{\phi F_L}\tilde{l}^* &= \gamma^* - \left(2 - \frac{1-\alpha}{\alpha}\chi\right)\nu^* - (1-\alpha)z^*. && (68)
\end{aligned}$$

These four equations are readily solved for a unique steady state. Eliminating $\chi/(\phi F_L)\tilde{l}^*$ from (66) and (67) and (67) and (68), respectively gives:

$$\begin{aligned}
\frac{n}{\alpha} + \nu^* &= \alpha z^* && (69)\\
\left(\frac{1-\alpha}{\alpha} - \frac{1}{\chi}\right)n + \nu^* &= \gamma^* - (1-\alpha)z^*. && (70)
\end{aligned}$$

Equation (65) can be used to eliminate γ^* from (70):

$$\left(\frac{1}{\alpha} - \frac{1}{\sigma} - \frac{1}{\chi}\right)n + \nu^* + \alpha\left(\frac{\alpha}{\sigma} - 1\right)z^* = \frac{\rho}{\sigma}. \tag{71}$$

Let $\Delta = (\sigma-1)n/\chi + \rho$. Combining (69) and (71) to derive ν^* and z^*, and using (67) and (68) gives (see the detailed derivation in Appendix 6.1)

$$\nu^* = \frac{1}{\alpha}\left(\Delta + \frac{1}{\chi}n\right) \tag{72}$$

$$z^* = \frac{1}{\alpha^2}\left(\Delta + \frac{1+\chi}{\chi}n\right) \tag{73}$$

$$\gamma^* = \frac{1}{\alpha^2}\left[\Delta + \frac{1+\chi}{\chi}(1-\alpha^2)n\right] \tag{74}$$

$$\tilde{l}^* = \frac{\phi F_L}{\alpha}\left[\Delta + \frac{1+\alpha}{\chi}n\right]. \tag{75}$$

Since $\tilde{l}^* = l^*\phi F_L$,

$$l^* = \frac{1}{\alpha}\left(\Delta + \frac{1+\alpha}{\chi}n\right). \tag{76}$$

The steady state in (72)-(74) and (76) is the steady state in the Jones (1995) model with $a \equiv \phi F_L$ (where "a" in Jones (1995) is the inverse of the R&D productivity). The present modeling of firm heterogeneity and R&D implies that entry costs and firm heterogeneity exclusively affect the productivity of R&E, and hence $\tilde{l} = L/A^\chi$, in the long-run.

Arnold (2006)'s findings on the dynamics of the Jones (1995) model are thus robust to our extensions. In particular, $\Delta > 0$ is sufficient to ensure that all stationary variables are positive, utility is bounded ($(1-\sigma)\dot{c}/c - \rho < 0$ since $\dot{c}/c = n/\chi$), and that the transversality condition holds, so that a steady state exists if and only if $\Delta > 0$.

Using the steady state values we find that this "independence" result carries over to the allocation of labor.

Result 8 (BGP labor shares)**.** *In steady state, the allocation of labor between R&E and production is independent of F_L.*

This observation is easily verified by combining the condition for optimal labor in production in (16), the free entry condition in aggregate terms in (42), and ν^*:

$$\nu^* = \frac{(1-\alpha)Y}{A\phi v_L} = \frac{wL_Y}{wA^\chi \phi F_L} = \frac{L_Y}{A^\chi \phi F_L} = \frac{L_Y}{L}l^*,$$

and therefore,[42]

$$\frac{L_Y}{L} = \frac{\nu^*}{l^*} = \frac{\Delta + \frac{n}{\chi}}{\Delta + \frac{1+\alpha}{\chi}n}. \tag{77}$$

[42]Explicitly, (77) states $\frac{L_Y}{L} = \frac{\sigma n + \chi\rho}{(\sigma+\alpha)n + \chi\rho}$.

As $\alpha > 0$, and $\Delta + n/\chi > 0$, we have $0 < L_Y/L < 1$. From this observation and the labor market clearing condition (7), we directly infer that $0 < L_A/L < 1$ and $L_A/L = 1 - \frac{\nu^*}{l^*}$ is given irrespective of the level of F_L.[43]

Similarly, using (3), (11), (24), and $G(b_L^*) = F_B/[(\phi - 1)F_L]$, in steady state,

$$\left(\frac{L_B}{L}\right)^* = \frac{n}{\chi}\frac{A^\chi}{L}(\phi - 1)F_L \tag{78}$$

$$\left(\frac{L_E}{L}\right)^* = \frac{n}{\chi}\frac{A^\chi}{L}F_L. \tag{79}$$

Since the labor shares are independent of the barriers to entry, the total labor income is also independent of the entry costs. Due to the Cobb-Douglas production of output we have $wL_Y = (1-\alpha)Y$, see 13, and hence

$$\frac{wL}{Y} = (1-\alpha)\frac{L}{L_Y} = (1-\alpha)\frac{\nu^*}{l^*}.$$

No reallocation between factor incomes. Do entry costs affect the aggregate distribution of wage and capital income in the steady state? Using the definition of z in $wL_Y = (1-\alpha)Y$ from the last paragraph, $wL_Y = (1-\alpha)r^*K/\alpha^2$ and hence

$$\frac{wL}{rK} = \frac{(1-\alpha)l^*}{\alpha^2\nu^*}$$

is independent of barriers to entry.

Having characterized the equilibrium in the closed economy, we now turn to the open economy.

3 Trade

We include characteristic features of international trade in the simplest possible way. Consider a world of two economies, each one as described in the previous section. The two economies have identical preferences, technologies, production structures, and identical capital and labor endowments. Only intermediate goods are traded internationally.[44] The free flow of intermediates between the two countries is hampered by marginal trading costs and TBTs.

[43] $\frac{L_A}{L} = \frac{\alpha n}{[(\sigma+\alpha)n+\rho\chi]}$.

[44] Including trade in the final good would allow imports of new physical capital.

TBTs. Empirically, TBTs remain important obstacles between developed countries despite various rounds of free trade negotiations. Importantly, TBTs are pure trading costs, and as such should be interpreted distinctly from the local market entry costs. TBTs are fixed costs associated with the entry of firms into the export market and account for country specific product/production standards/regulations, additional certification procedures, or additional bureaucratic burdens that make it harder for foreign firms to supply the domestic market than for their local competitors.[45] To capture the relative disadvantage for foreign firms, we follow Melitz (2003) and assume that foreign firms face higher entry costs when entering the export market than local firms that enter that same market. Due to symmetry, exporting thus comes at higher fixed costs than producing for the local market from a domestic firm's point of view. Hence, there is another cutoff productivity, $1/b_E$, for exporting. In the presence of TBTs, b_E is lower than b_L, so that the equilibrium productivity pattern in the local and the export market matches the empirical regularity that the bulk of firms sells only locally and only the most productive firms export. As an aside, the empirical trade literature has also clarified that there are no feedback effects from exporting to a firm's productivity (Bernard and Jensen, 1999, and Bernard, Jensen and Schott, 2006). Hence, the fact that input coefficients remain constant is in line with recent empirical evidence. Returning to the model, we account for the different sources of TBTs and iceberg costs by using different "factor intensities" for the two types of trade barriers. With respect to the fixed export costs, we adapt the modeling in the literature and assume that fixed export costs are wage costs.

Iceberg costs. The variable trading costs are modeled as Samelson-type iceberg costs and as such decrease the productivity in the production of exported units. Since intermediate goods are manufactured using physical capital, variable trading costs constitute capital costs to the firms.

In what follows, we describe the additional assumptions for the open economy and then deduce the adjusted optimality conditions.

[45] Roberts and Tybout (1997) validate empirically that the sunk cost associated with exporting are of a substantial magnitude for exporting firms.

3.1 Open economy

3.1.1 Technologies

Output. We distinguish local and export market variables of intermediate firms by a subscript i, $i \in \{L, E\}$. To avoid additional complexity, we focus on the case where there is no overlap between local and foreign varieties at the point in time when trade liberalization occurs (cf. Tang and Wälde, 2001). That is, if foreign firms decide to export, local producers of output are able to employ a higher number of intermediate inputs, thereby exploiting increased gains from specialization. We denote the worldwide "number" of different existing intermediate goods by $\tilde{A} > A_L$. Without loss of generality, let the intermediate goods index be such that $j \in [0, A_L]$ indicates locally produced intermediate goods, while $j \in (A_L, \tilde{A}]$ refers to imported varieties. Aggregate output in each economy is given by

$$Y = L_Y^{1-\alpha} \int_0^{\tilde{A}} x_i(j)^\alpha dj, \tag{80}$$

where $x_i(j) = x_L(j)$ for $j \leq A_L$ represents the input quantity of a locally produced good and $x_i(j) = x_E(j)$ for $j > A_L$ is the input quantity of an imported intermediate good.

Intermediates. When exported, $\tau \geq 1$ units of an intermediate good must be shipped for every unit that arrives (exporters get paid for the arriving units). From an exporting firm's perspective, producing one (actually sold) unit for the foreign market thus ties up τ times more resources than producing the same unit for the local market:

$$x_i(j) = \frac{k_i(j)}{\tau_i b(j)}, \tag{81}$$

where here and in what follows $\tau_i = \tau$ if goods are exported ($i = E$), and $\tau_i = 1$ if goods are sold locally ($i = L$). Iceberg costs therefore imply that the productivity of a firm depends on the destination of the manufactured output. The productivity in the production of exports thereby decreases linearly in the iceberg costs. To simplify the exposition, we treat the marginal trade costs as a technology such that they do not yield any income.[46]

[46] See Matsuyama (2007) for a theory of factor biased trading costs.

3.1.2 Markets

Barriers to trade. Launching an export business with a newly discovered intermediate good requires the entrant to hire $A_L^{\chi-1}F_E$ "entry workers", where $F_E \geq F_L$ so that $T \equiv F_E/F_L \geq 1$. T (a mnemonic for TBTs) measures how much harder it is for a foreign firm to enter the local market compared to a domestic firm. If $T > 1$, additional profit opportunities from exporting accrue only to the most productive firms which can profitably afford to sink the foreign market entry costs. Below, incremental trade liberalization is modelled as a decrease in either transportation costs, τ, or TBTs, T. In the case of TBTs, we formally consider the comparative statics with respect to F_E (evaluated at $F_E \geq F_L$).

Profitable market entry. Firms base the decision to export on the same forward-looking investment calculus as the decision to enter the domestic market in autarky. Given its productivity, each firm knows its future profits in either market and decides to enter a market only if the present value of its profits in that market exceeds the entry costs to that market. We denote by $1/b_L$ and $1/b_E$ the lowest productivity levels that allow firms to operate profitably in the local and the export market, respectively. The presence of TBTs (i.e. $F_E > F_L$) implies that $b_E < b_L$ (we will see below that profits increase monotonically in productivity), so that only the most productive firms export. Hence, $x_E(j)$ in (80) is zero for some $j > A_L$.[47] While entering the domestic market is less involved for local firms, it still takes costly resources and hence there is a minimum productivity requirement for active firms (the lower bound on F_L in (PA1) ensures that $b_L < b_0$ in equilibrium so that the minimum productivity requirement is binding for some firms). The least productive firms, which do not meet the minimum productivity requirement, exit immediately. Firms with input coefficients $b_L \geq b > b_E$ sell exclusively in their home market and the most productive firms with $b \leq b_E$ sell both in the local market and export.

Market entry and research again use the same production technology and we define its productivity to include the effects of international knowledge spillovers.[48] For further reference

[47]By the definition of A_L as the number of actually active firms, all $x_L(j)$ are positive for $j \leq A_L$.

[48]Consider the Jones (1995) R&D production function from the closed economy, augmented by international

let

$$v_i(j) \equiv \int_t^\infty e^{-\bar{r}(s-t)}\pi_i(j)ds \tag{82}$$

be the present value of operating profits $\pi_i(j)$ of firm j in market i.

Since firms differ only in terms of their productivities, profit maximization together with (82) implies that $v_i(j; b(j)) = v_i(j'; b(j'))$ if and only if $b(j) = b(j')$. With a slight abuse of notation, we can therefore drop the firm index and equivalently state the present value of profits of firm j in market i as a function of the firm's input coefficient (and other variables), i.e.

$$v_i(b,\cdot) \equiv \int_t^\infty e^{-\bar{r}(s-t)}\pi_i(b,\cdot)ds \tag{83}$$

where $\pi_i(b,\cdot)$ is the operating profit of a firm with input coefficient b in market i. Differentiating (83) with respect to time t, the definition of $v_i(b,\cdot)$ implies

$$rv_i(b,\cdot) = \pi_i(b,\cdot) + \dot{v}_i(b,\cdot). \tag{84}$$

Entry into R&D. Including the additional profit opportunities net of entry costs for productivity types $b \leq b_E$, free entry into R&D in the open economy requires

$$\int_0^{b_0} \max\left[v_L(b,\cdot) - A_L^{\chi-1}wF_L, 0\right]dG(b) + \int_0^{b_0} \max\left[v_E(b,\cdot) - A_L^{\chi-1}wF_E, 0\right]dG(b) = A_L^{\chi-1}wF_B \tag{85}$$

if $\dot{B}_L > 0$. In equilibrium, the innovation costs have to equal the expected market value of a newly discovered blueprint, i.e. the sum of the expected operating values net of entry costs in both the local and the foreign markets. Expectations are taken with respect to productivity, accounting for the fact that only sufficiently productive blueprints sell at all and that only

---

knowledge spillovers:

$$\dot{B} = \frac{(A_L + \sigma A_F)^{1-\chi}L_A}{\tilde{F}_B},$$

where A_F is the knowledge stock in the foreign country, $\sigma \geq 0$ measures the intenisty of the across-the-border spillovers, and $\tilde{F}_B$ is an exogenous productivity parameter. With symmetric countries it holds that $A_F = A_L$ and hence $A_L + \sigma A_F = (1+\sigma)A_L$. We let $F_B \equiv \tilde{F}_B/(1+\sigma)^{1-\chi}$ to maintain

$$\dot{A}_L = \frac{A_L^{1-\chi}L_A}{F_B}$$

as in the closed economy.

blueprints with a high productivity allow for additional profits in the export market. We show below that market values are monotonely decreasing in b so that the cutoffs b_L and b_E, i.e. the productivities that yield zero in the squared brackets in (85), are unique.[49] In particular, the cutoff associated with exporting, b_E, is determined via

$$v_E(b_E, \cdot) = wA_L^{\chi-1}F_E, \tag{86}$$

whereas the minimum productivity requirement for active firms, $1/b_L$, is implicitly given by

$$v_L(b_L, \cdot) = wA_L^{\chi-1}F_L. \tag{87}$$

In view of (86) and (87), the free entry condition in (85) equivalently reads

$$\int_0^{b_L} \left[v_L(b, \cdot) - A_L^{\chi-1}wF_L\right] dG(b) + \int_0^{b_E} \left[v_E(b, \cdot) - A_L^{\chi-1}wF_E\right] dG(b) = A_L^{\chi-1}wF_B. \tag{88}$$

No imitation or "footloose" production. Two simplifying assumptions ensure that firms have no means to avoid the trading costs. First, firms are not able to form multinational companies or issue production licenses, i.e. there is no "footloose" production. Second, transportation costs are lower than the cost of patent infringement, so that imitation and limit pricing by foreign firms is not profitable.

3.2 Equilibrium

We proceed by deriving the equilibrium for given cutoff levels, then aggregate, and use the results to determine the cutoff productivity levels.

3.2.1 Optimality conditions

Households and firms. Optimal consumption is not affected by the degree of trade openness. The demand for intermediates from the final good sector in (14) now applies to all $j \in [0, \tilde{A}]$ and the profit-maximizing monopoly price in (15) refers to the price of selling locally, $p_L(j) = rb(j)/\alpha$. When exporting, the monopolists charge the profit maximizing price

[49] Under our parameter assumptions, both cutoffs also exist within the support of the equilibrium distribution of active firms' productivities.

$p_E(j) = \tau r b(j)/\alpha = \tau p_L(j)$. Using these pricing rules, the demand for variety j in market i equals

$$x_i(j) = \alpha^{2\epsilon} \left[r\tau_i b(j) \right]^{-\epsilon} L_Y. \tag{89}$$

If firm j is active in market i, it receives equilibrium revenues

$$R_i(j) = \alpha^{2\epsilon-1} \left[r\tau_i b(j) \right]^{1-\epsilon} L_Y, \tag{90}$$

and earns profits $\pi_i(j) = (1-\alpha)R_i(j)$ in that market.From the fact that $p_i(j) = p_i(j')$ and $x_i(j) = x_i(j')$ if and only if $b(j) = b(j')$, equilibrium prices and quantities can equivalently be phrased in terms of productivities, i.e.

$$p_i(b,\cdot) = \frac{\tau_i r b}{\alpha},\ x_i(b,\cdot) = \alpha^{2\epsilon} \left[r\tau_i b \right]^{-\epsilon} L_Y.$$

The same is true for revenue and profits,

$$R_i(j) = \alpha^{2\epsilon-1} \left[r\tau_i b(j) \right]^{1-\epsilon} L_Y,\ \mathrm{dm}\ \pi_i\left(b,\cdot\right)$$

and was already used for the present value of profits as defined above, see (83).

A simple connection between the cutoffs. Due to the symmetry assumption, the relation of the two cutoff productivities is easily derived as follows. Combining the local and the foreign market entry condition, (86) and (87), gives $v_E(b_E,\cdot)/v_L(b_L,\cdot) = T$. For a given b, $v_E(b,\cdot)$ and $v_L(b,\cdot)$ differ only because of the marginal trade costs. Since $\hat{v}_i(b,\cdot) = \hat{v}(\cdot)$ from (84), $\pi_E(b,\cdot) = \pi_L(\tau b,\cdot)$ implies $v_E(b,\cdot) = v_L(\tau b,\cdot)$. We know from the closed economy that the ratio of any two firms' market values only depends on their input coefficients (and ϵ, see (22)). In this symmetric setup, the two cutoff input coefficients are therefore exclusively related by variable and fixed barriers to trade:

$$\frac{b_E}{b_L} = \psi, \quad 0 > \psi \equiv \tau^{-1} T^{-\frac{1}{\epsilon-1}} \geq 1. \tag{91}$$

ψ is an inverse measure of the real barriers to trade and measures the economies' openness. Under free trade, $\tau = T = 1$ and $\psi = 1$. The more restricted trade is (i.e. the larger τ and T), the smaller is ψ. Autarky corresponds to $\psi \to 0$.

Marginal trade costs drive a wedge between the minimum productivity requirement for the local and the export market, and TBTs have a more severe impact the larger ϵ. A high elasticity of substitution in the production of output depresses prices, profits, and hence market values so that firms must be more productive to cover the entry costs.[50]

3.2.2 Aggregation

Output. Making use of the convention that goods $j \in [0, A_L]$ refer to locally produced intermediates while $j \in [A_L, \tilde{A}]$ indicate imported varieties, aggregate output is

$$Y = L_Y^{1-\alpha} \int_0^{\tilde{A}} x(j)^\alpha dj = L_Y \left[\int_0^{A_L} x_L(j)^\alpha dj + \int_{A_L}^{\tilde{A}} x_E(j)^\alpha dj \right].$$

Switching from firms/goods j to productivities b, the term in squared brackets becomes $A_L \int_0^{b_L} x_L(b)^\alpha d\mu(b) + A_L^\diamondsuit \int_0^{\psi b_L^\diamondsuit} x_E(b)^\alpha d\mu^\diamondsuit(b)$, where the diamond indicates foreign values, and $b_E^\diamondsuit = \psi b_L^\diamondsuit$ was used (see (91)). We again solve the model in terms of the local cutoff productivity and hence aggregate to get expressions in terms of b_L. Due to symmetry, $A_L = A_L^\diamondsuit$, $b_L = b_L^\diamondsuit$ (s.t. $\mu^\diamondsuit(b) = \mu(b)$) and using (89) and (91), we have

$$Y = L_Y^{1-\alpha} \left[A_L \alpha^{2\alpha\epsilon} r^{-\alpha\epsilon} L_Y^\alpha \left(\int_0^{b_L} b^{-\alpha\epsilon} d\mu(b) + \tau^{-\alpha\epsilon} \int_0^{\psi b_L} b^{-\alpha\epsilon} d\mu(b) \right) \right].$$

Applying the Pareto specification and integrating yields

$$\begin{aligned} Y &= L_Y^{1-\alpha} \left[\frac{A_L \alpha^{2\alpha\epsilon} r^{-\alpha\epsilon} L_Y^\alpha \theta}{b_L^\theta} \left(\int_0^{b_L} b^{\theta-\epsilon} db + \tau^{1-\epsilon} \int_0^{\psi b_L} b^{\theta-\epsilon} db \right) \right] \\ &= L_Y^{1-\alpha} \left[\phi A_L \alpha^{2\alpha\epsilon} r^{-\alpha\epsilon} L_Y^\alpha \left(b_L^{1-\epsilon} + \tau^{1-\epsilon} \psi^{\theta-\epsilon+1} b_L^{1-\epsilon} \right) \right]. \end{aligned}$$

Using (89) again and recognizing $-\alpha\epsilon = 1 - \epsilon$, aggregate output can be expressed as

$$Y = \phi A_L L_Y^{1-\alpha} x(b_L, \cdot)^\alpha \left(1 + \tau^{1-\epsilon} \psi^{\theta-\epsilon+1} \right).$$

After collecting the parameters in the last term, we get

$$Y = \phi \Psi A_L L_Y^{1-\alpha} x(b_L, \cdot)^\alpha, \tag{92}$$

where

$$1 \leq \Psi \equiv 1 + \psi^\theta T \leq 2, \quad \partial\Psi/\partial\tau < 0, \ \partial\Psi/\partial T < 0.$$

[50] Note again that α also measures the capital intensity in the final good production and the gains from specialization.

Relative to the closed economy (compare (37)), the only difference is the additional parameter Ψ. Prohibitive trade costs imply $\Psi = 1$ (autarky), and free trade ($T = \tau = 1$) corresponds to $\Psi = 2$.

Capital. The capital stock is analogously derived as follows. From

$$K = \int_0^{A_L} k(j)dj = \int_0^{A_L} b(j)x(j)dj,$$

where $b(j)$ and $x(j)$ refer to the total output of an intermediate firm (i.e. output sold in the local market and, if applicable, output sold in the export market) and the destination dependent productivity, respectively. In terms of productivities,

$$K = A_L \int_0^{b_L} bx_L(j)d\mu(b) + A_L \int_0^{b_E} b\tau x_E(j)d\mu(b).$$

Integrating over the distribution of firms' productivity types we get

$$\begin{aligned} K &= A_L\alpha^{2\epsilon}r^{-\epsilon}L_Y \left[\int_0^{b_L} b^{1-\epsilon}d\mu(b) + \int_0^{b_E} (\tau b)^{1-\epsilon}d\mu(b)\right] \\ &= \frac{A_L\alpha^{2\epsilon}r^{-\epsilon}L_Y\theta}{b_L^\theta}\left[\int_0^{b_L} b^{\theta-\epsilon}db + \int_0^{\psi b_L} \tau^{1-\epsilon}b^{\theta-\epsilon}db\right] \\ &= \phi A_L\alpha^{2\epsilon}(rb_L)^{-\epsilon}L_Y b_L(1+\tau^{1-\epsilon}\psi^{\theta-\epsilon+1}). \end{aligned}$$

The last term is again $1 + \tau^{1-\epsilon}\tau^{\epsilon-1-\theta}T^{-\frac{\theta}{\epsilon-1}+1} = \Psi$, such that

$$K = \phi\Psi A_L b_L x(b_L, \cdot). \tag{93}$$

After replacing $x(b_L, \cdot)^\alpha = K^\alpha(\phi\Psi A_L b_L)^{-\alpha}$, aggregate output in the open economy with costly trade is

$$Y = (\phi\Psi A_L L_Y)^{1-\alpha}\left(\frac{K}{b_L}\right)^\alpha. \tag{94}$$

Relative to the closed economy, the only difference is again the parameter Ψ.

Aggregate revenues, profits, and market values. From (80) and (90), the total payments for intermediate goods in each economy equal

$$\int_0^{\tilde{A}} R(j)dj = \int_0^{A_L} R_L(j) + \int_{A_L}^{\tilde{A}} R_E(j)dj.$$

Trade balance gives that the revenues of foreign producers (the second term) equal the revenues of domestic firms from exporting,

$$\int_{\tilde{A}-A_L}^{\tilde{A}} R_E(j)dj = \int_0^{A_L} R_E(j)dj$$

where $R_E(j) = 0$ for all firms with $b(j) > b_E$. Accordingly,

$$\int_0^{\tilde{A}} R(j)dj = \int_0^{A_L} R(j)dj,$$

where $R(j)$ denotes the total revenues accruing to firm j, i.e. $R_L(b, \cdot)$ for $b_L \geq b > b_E$, and $R_L(b, \cdot) + R_E(b, \cdot)$ for $b \leq b_E$. In terms of productivities,

$$\int_0^{A_L} R(j)dj = A_L \int_0^{b_L} p_L(b, \cdot)x_L(b, \cdot)d\mu(b) + A_L \int_0^{b_E} p_E(b, \cdot)x_E(b, \cdot)d\mu(b).$$

Using the pricing rules,[51]

$$\int_0^{A_L} R(j)dj = A_L \alpha^{2\epsilon-1} r^{1-\epsilon} L_Y \left[\int_0^{b_L} b^{1-\epsilon} d\mu(b) + \int_0^{b_E} (\tau b)^{1-\epsilon} d\mu(b) \right].$$

Applying the Pareto specification, and integrating again, we have

$$\int_0^{A_L} R(j)dj = A_L \alpha^{2\epsilon-1} (r b_L)^{1-\epsilon} L_Y \phi \left(1 + \tau^{1-\epsilon} \psi^{\theta-\epsilon+1}\right).$$

Recognizing that the last term equals again Ψ, after using (90), aggregate revenues are

$$\int_0^{A_L} R(j)dj = \Psi \phi A_L R_L(b_L, \cdot).$$

Since profits are a fraction $1 - \alpha$ of revenues,

$$\int_0^{A_L} \pi(j)dj = \Psi \phi A_L \pi_L(b_L, \cdot),$$

[51]We could take the following shortcut here: replacing the term in squared brackets with the expression from the derivation of aggregate output, the revenues of local producers amount to

$$\int_0^{A_L} R(j)dj = Y L_Y^{\alpha-1} A_L^{-1} \alpha^{-2\alpha\epsilon} r^{\alpha\epsilon} L_Y^{-\alpha} A_L \alpha^{2\epsilon-1} r^{1-\epsilon} L_Y = \alpha Y.$$

Since aggregate profits are a fraction $1 - \alpha$ of aggregate revenues, aggregate profits are

$$\int_0^{A_L} \pi(j)dj = \alpha(1 - \alpha)Y.$$

In an equilibrium with balanced growth, using (84), we have

$$\int_0^{A_L} v(j)dj = \alpha(1 - \alpha)Y/(r - \hat{v}).$$

This directly yields the open economy ν.

so that, along a balanced growth path, (84) implies

$$\int_0^{A_L} v(j)dj = \Psi\phi A_L v_L(b_L,\cdot).$$

Accounting for costly international trade simply adds the factor $\Psi \leq 2$. The average firm value is $\int_0^{A_L} v(j)dj/A_L = \Psi\phi v_L(b_L,\cdot)$. In the absence of trade, $\Psi = 1$, and we are back in the closed economy where ϕ relates the value of firms with the cutoff productivity to the average value of firms in the local market. If trade is costless, the average firm value is $\int_0^{A_L} v(j)dj/A_L = 2\phi v_L(b_L,\cdot)$. The aggregated average firm value of producers that are only selling to the local market was derived in the closed economy and equals

$$A_L \int_0^{b_L} v_L(b,\cdot)d\mu(b) = \phi v_L(b_L,\cdot).$$

Entry into R&D for given cutoffs. Equation (91) is the key in solving the open economy model in the same block recursive manner as the autarky model.

$$\int_0^{b_L} \left[v_L(b,\cdot) - A_L^{\chi-1}wF_L\right] dG(b) + \int_0^{b_E} \left[v_E(b,\cdot) - A_L^{\chi-1}wF_E\right] dG(b) = A_L^{\chi-1}wF_B. \tag{95}$$

Splitting up the integral terms, the free entry condition in (88) becomes

$$\int_0^{b_L} v_L(b,\cdot)dG(b) - A_L^{\chi-1}wF_L \int_0^{b_L} dG(b) + \int_0^{b_E} v_E(b,\cdot)dG(b) - A_L^{\chi-1}wF_E \int_0^{b_E} dG(b) = A_L^{\chi-1}wF_B, \tag{96}$$

or, equivalently,

$$\int_0^{b_L} v_L(b,\cdot)dG(b) + \int_0^{b_E} v_E(b,\cdot)dG(b) = A_L^{\chi-1}w\left[F_B + G(b_L)F_L + G(b_E)F_E\right]$$

After dividing by $G(b_L)$ and recognizing from the definition of $\mu(b)$ in (29) that

$$\frac{dG(b)}{G(b_L)} = \mu'(b)\,db = d\mu(b),$$

we have

$$\int_0^{b_L} v_L(b,\cdot)d\mu(b) + \int_0^{b_E} v_E(b,\cdot)d\mu(b) = A^{\chi-1}w\left[\frac{F_B}{G(b_L)} + F_L + F_E\frac{G(b_E)}{G(b_L)}\right]. \tag{97}$$

The term in squared brackets on the right-hand side is, like $F^a(b_L)$ in the closed economy, the quantity of labor necessary to invent and market a new intermediate good in the absence of knowledge spillovers ($\chi = 1$). Using (91), it can be expressed as a function of the local cutoff only. Substituting for b_E with (91) and using the Pareto specification,

$$G(b_E)/G(b_L) = \psi^\theta.$$

Hence,

$$\frac{F_B}{G(b_L)} + F_L + F_E\frac{G(b_E)}{G(b_L)} = \frac{F_B}{G(b_L)} + F_L + \theta^{-\theta} F_E^{1-\frac{\theta}{\epsilon-1}} F_L^{\frac{\theta}{\epsilon-1}} = \frac{F_B}{G(b_L)} + F_L\left[1 + \left(\frac{F_E}{F_L}\right)^{1-\frac{\theta}{\epsilon-1}} \tau^{-\theta}\right].$$

To avoid cumbersome expressions, we keep in mind that T is contained in ψ also, but collect parameters so that the last term on the right-hand side becomes

$$\frac{F_B}{G(b_L)} + F_L\left[1 + \left(\frac{F_E}{F_L}\right)^{1-\frac{\theta}{\epsilon-1}} \tau^{-\theta}\right] = \frac{F_B}{G(b_L)} + F_L\left(1 + \psi^\theta T\right) = \frac{F_B}{G(b_L) + \Psi F_L} \equiv F\left(b_L\right). \quad (98)$$

The left-hand side of (97) can be expressed in similar terms. Using $\hat{v}_i(b, \cdot) = \hat{v}(\cdot)$, $d\mu = \theta b^{\theta-1}/b_L^\theta db$, and (91), the average value of a usable blueprint becomes

$$\frac{(1-\alpha)\alpha^{2\epsilon-1}r^{1-\epsilon}L_Y\theta}{(r-\hat{v})b_L^\theta}\left(\int_0^{b_L} b^{\theta-\epsilon}db + \tau^{1-\epsilon}\int_0^{\psi b_L} b^{\theta-\epsilon}db\right).$$

Integrating and rearranging terms yields

$$\frac{(1-\alpha)\alpha^{2\epsilon-1}r^{1-\epsilon}L_Y\theta}{(r-\hat{v})b_L^\theta}\left[\frac{b_L^{\theta-\epsilon+1}}{\theta-\epsilon+1} + \tau^{1-\epsilon}\frac{(\psi b_L)^{\theta-\epsilon+1}}{\theta-\epsilon+1}\right],$$

or equivalently

$$\frac{\phi(1-\alpha)\alpha^{2\epsilon-1}(rb_L)^{1-\epsilon}L_Y\left(1+\tau^{1-\epsilon}\psi^{\theta-\epsilon+1}\right)}{r-\hat{v}}.$$

From (89), this equals $\phi v_L + \psi^{\theta-(\epsilon-1)}\phi v_E$. Costly trade lowers the value of exporting relative to local sales ($\psi \in (0, 1]$ and $\theta > \epsilon - 1$). Like in the closed economy, ϕ relates the average market value of exclusively locally selling firms to the market value of firms with the local market cutoff productivity. Since we aim at aggregating in terms of the local market cutoff, we rewrite the average value of a usable blueprint as

$$\phi v_L\left(1+\tau^{1-\epsilon}\psi^{\theta-\epsilon+1}\right) = \phi v_L\left[1+\tau^{-\theta}T^{-\left(\frac{\theta}{\epsilon-1}-1\right)}\right] = \Psi\phi v_L.$$

Again for expositional convenience, we collect parameters in the last equation to have

$$\phi v_L \left(1 + \psi^{\theta} T\right). \tag{99}$$

Using (98) and (99), the free entry into innovation condition in terms of the local market cutoff reads

$$\Psi \phi v_L = w A_L^{\chi - 1} F(b_L). \tag{100}$$

We now solve the model for its steady-state equilibrium. In the following section, we provide additional economic intuition and discuss the reallocation and incentive effects induced by international trade.

3.2.3 Equilibrium

In view of the close and block-recursive relation of b_L and b_E in (91), the definition of an equilibrium in the closed economy carries over to the open economy (including b_E, the additional equation is (91)). The steady state growth rates are the same as in the closed economy, and $\hat{b}_E = 0$ follows from (91). The cutoffs are again instantaneously fixed.

Equilibrium cutoffs. Combining the entry conditions for the local product market and for R&D in (87) and (100) again pins down the local productivity cutoff b_L^*, and via (91) b_E^* also. In equilibrium, the share of sufficiently productive blueprints is

$$G(b_L^*) = \frac{F_B}{(\phi - 1) F_L \Psi}, \tag{101}$$

and the implied cutoffs are

$$\begin{aligned} b_L^* &= \left[\frac{F_B}{(\phi - 1) F_L \Psi}\right]^{\frac{1}{\theta}} b_0, \\ b_E^* &= \psi \left[\frac{F_B}{(\phi - 1) F_L \Psi}\right]^{\frac{1}{\theta}} b_0 \quad (\geq b_L^*). \end{aligned}$$

Inserting the definition of Ψ, the local market cutoff is

$$b_L^* = \left[\frac{F_B}{F_L(\phi - 1)}\right]^{\frac{1}{\theta}} \frac{b_0}{\left(1 + \tau^{-\theta} T^{1 - \frac{\theta}{\epsilon - 1}}\right)^{-\frac{1}{\theta}}}. \tag{102}$$

The equilibrium cutoff associated with exporting is

$$b_E^* = \left[\frac{F_B}{F_L(\phi - 1)}\right]^{\frac{1}{\theta}} \frac{b_0 \tau^{-1} T^{-\frac{1}{\epsilon-1}}}{\left(1 + \tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}}. \tag{103}$$

Using (101), the average labor requirement for a usable blueprint in the absence of knowledge spillovers amounts to[52]

$$F(b_L^*) = \frac{F_B}{G(b_L^*)} + \Psi F_L = \Psi \phi F_L. \tag{104}$$

3.2.4 Dynamic Equilibrium

Laws of motions and steady state. Turning to the dynamic equilibrium, the same transformed variables as in the closed economy apply. The integral in the denominator of ν now goes from 0 to A_L, and $\nu = (1-\alpha)Y / \int_0^{A_L} v_i(b, \cdot) d\mu(b_L^*) = (1-\alpha)Y/[\Psi \phi A_L v_L(b_L^*)]$. Just like in the closed economy, $rK/\alpha{=}\alpha Y$, or $r = \alpha^2 z$. Also, using the adjusted definition of ν, $\dot{K}/K = z - \gamma$ and $\dot{v}/v = \alpha(\alpha z - \nu)$ continue to hold. The innovation costs change from $F^a(b_L^*) w A^{\chi-1} = \phi F_L w A^{\chi-1}$ to $F(b_L^*) w A^{\chi-1} = \Psi \phi F_L w A_L^{\chi-1}$, and this only affects the law of motion of A_L in terms of $\tilde{l}$: $\dot{A}_L/A_L = \tilde{l}/(\Psi \phi F_L) - \nu$. Accordingly, the laws of motion for γ and z are still given by (53) and (60), where ν in (60) refers to the adjusted definition from above. The law of motion for $\tilde{l}$ becomes

$$\dot{\tilde{l}} = \tilde{l}\left[n - \chi\left(\frac{\tilde{l}}{\Psi \phi F_L} - \nu\right)\right].$$

Including the $\dot{A}_L/A_L$, the law of motion for ν reads

$$\dot{\nu} = \nu\left[\left(\frac{1-\alpha}{\alpha}\chi - 1\right)\frac{\tilde{l}}{\Psi \phi F_L} + \left(2 - \frac{1-\alpha}{\alpha}\chi\right)\nu + (1-\alpha)z - \gamma\right].$$

Accordingly, the steady state in the open economy is the steady state in the closed economy in terms of ν^*, z^*, γ^*, and l^*. Ψ only affects the resources in R&E, and

$$\tilde{l}^* = \frac{\Psi \phi F_L}{\chi}\left[\Delta + \frac{1+\alpha}{\chi} n\right].$$

[52]For free entry into R&D to be in line with the local market entry condition, it must be that the impact of trade on the average firm value (relative to the local cutoff productivity type firm value) is equal to the impact of trade on the average development costs (relative to the entry costs), i.e. Ψ must drop out in the free entry into R&D condition, compare (100) using (104).

4 Trade liberalization

To ease the exposition, suppose without loss of generality in the following analysis that $b_0 = 1$ and $F_B = F_L(\phi - 1)$. Then,

$$b_L^* = \frac{1}{\left(1 + \tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}}. \tag{105}$$

$$b_E^* = \frac{\tau^{-1} T^{-\frac{1}{\epsilon-1}}}{\left(1 + \tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}}. \tag{106}$$

With free trade, $\tau = T = 1$, and all firms are exporting ($b_E = b_L$). In this case, output in both economies is produced using $\tilde{A} = 2A_L$ intermediate goods at any point in time.

We now consider policy induced changes in the barriers to trade and show the presence of a Melitz (2003)-type reallocation towards the more productive firms.

4.1 Cutoffs and industry reallocation

Differentiating (106) exemplarily with respect to τ yields

$$\frac{\partial b_E^*}{\partial \tau} = \frac{-\left(1 + \tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}} \tau^{-2} T^{-\frac{1}{\epsilon-1}} + \tau^{-1} T^{-\frac{1}{\epsilon-1}} \left(1 + \tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}-1} \tau^{-\theta-1} T^{1-\frac{\theta}{\epsilon-1}}}{[.]^2} < 0.$$

The derivative with respect to T has the same sign.[53] It is negative, implying that a decrease in both types of trade costs increases b_E^* and hence lowers the minimum productivity requirement necessary for profitable exporting.[54] Simple inspection of (105) shows that a reduction in the barriers to trade ($\tau \downarrow$, $T \downarrow$) has the opposite effect on the local cutoff:

$$\frac{\partial b_L^*}{\partial \tau} > 0, \quad \frac{\partial b_L^*}{\partial T} > 0.$$

A decrease in either τ or T raises the minimum productivity requirement for all firms. Taken together, trade liberalization allows more firms to export profitably (and implies an increase in the intensive margin), but at the same requires all newcomers to be more productive.

[53] More precisely, to analyze the effect of a decrease in TBTs, we take the derivative with respect to F_E and evaluate it at $F_E \geq F_L$.

[54] $sgn\left[-\tau^{-2} T^{-\frac{1}{\epsilon-1}} + \tau^{-1} T^{-\frac{1}{\epsilon-1}} \left(1 + \tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}\right)^{-1} \tau^{-\theta-1} T^{-1-\frac{\theta}{\epsilon-1}}\right] = sgn\left(\frac{\tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}}{1+\tau^{-\theta} T^{1-\frac{\theta}{\epsilon-1}}} - 1\right) = -1.$

Result 9 (Reallocation)**.** *Trade liberalization lowers the productivity requirement for exporting, but increases the minimum productivity requirement for newcomers.*

The implied reallocation of resources from less productive firms towards more productive firms is the same as in Melitz (without productivity growth) and BRN (with fully endogenous steady state growth and scale effects). Including capital in production, factor prices are irrelevant for the determination of the cutoffs as long as R&D and entry are conducted with identical production functions. Therefore, GS find exactly the same cutoffs.

4.2 Labor shares

A direct implication of identical steady state values in autarky and in the open economy is that the allocation of labor between production and R&E is not affected by the exposure to trade. Moreover, for the reasons explained in detail in the closed economy (compare Section 2.7), the allocation of labor between R&D and entry is not affected by a reduction in trade costs.

Result 10 (No impact on labor shares)**.** *The allocation of labor in R&D, market entry, and production is not affected by the exposure to trade.*

4.3 Incentive effects of trade liberalization

We have seen three channels by which trade openness affects the incentives to innovate. Firstly, trade liberalization increases the minimum productivity requirement for all firms and thereby raises the average discovery costs of newly invented varieties. Secondly, at the same time, the expected value of a usable blueprint increases because a reduction of TBTs lowers the entry costs for the export market. Thirdly, this innovation enhancing effect from an increase in the returns to successful R&D is reinforced by the fact that more blueprints can be used to launch a profitable export business.

Growth effects under fully endogenous growth. BRN study the growth effects of trade liberalization in a fully endogenous growth framework (with scale effects). They find that openness to trade is growth enhancing if and only if the expected sunk cost of R&E decrease (their Result 1, p. 10).

The sunk costs of R&E consist of the quantity of workers necessary to conduct R&E for marketable blueprints, magnified by the impact of spillovers, and the associated wage. BRN state that the actual labor requirement unambiguously increases if a country incrementally opens up to trade. The impact on the price for R&E depends on the exact specification of the engine of growth, but is likely to be positive. In the Grossman-Helpman specification (Grossman and Helpman, 1991, Ch.3), the net-effect permanently depresses growth.

Incentive effects under semi-endogenous growth. In our formulation, the impact on the labor requirement is the similar. The innovation enhancing reduction in the labor requirement for entry is offset by the increase in the labor requirement due to a lower local cutoff:[55]

$$F(b_L^*) = \Psi\phi F_L = \frac{\phi}{\phi - 1}\Psi F_B$$

increases as T and/or τ decrease (which increases Ψ). If trade where free, $F(b_L^*) = 2\phi F_L$.

GS study how trade affects the level of total productivity (which in their model coincides with per capita consumption), i.e. variety growth and the productivity in production. To do so, they compare per capita consumption along the steady-state path of two economies which exclusively differ in terms of the trade costs. Since the ratio is constant over time, they conclude that trade increases productivity if and only if the path associated with lower barriers to trade has higher per capita consumption. This conclusion depends on the strength of knowledge spillovers in R&D. If spillovers are sufficiently strong, trade liberalization retards productivity growth in the short run (and makes consumers worse off in the long-run).

5 Conclusion

Recap. We lay out a specific environment to investigate how trade affects endogenous R&D in a dynamic model with heterogeneous firms and costly trade. Focusing on trade in non-durable intermediate goods, we highlight important features of the production structure when firms use physical capital and showed that trade in intermediate goods and a careful introduction of capital in production does not alter previous findings on the reallocation of resources and the selection of firms. We further find that, albeit convenient, the assumption of identical

[55]The second equality follows by replacing F_L with its normalized value in terms of F_B.

production technologies in R&D and entry is particularly restrictive. In particular, the labor shares between innovation and market entry are fixed independently of the level of barriers to entry.

Avenues for future research. The present paper provides a framework for various robustness checks. In particular, including trade in the final good and trade in durables is a straightforward extension. More importantly, however, the environment described in this paper allows us to include physical capital and/or units of output as an input in the entry process. A second extension concerns the average R&D costs approach. Recall that inventing a blueprint is costly, but the productivity implied by the discovered blueprint was a random draw. Hence, there is no intentional investment in productivity, and high productivity types come "for free". Exploring the trade-off between high productivities and the number of usable blueprints at the level of an individual researcher and thereby allowing for purposive investment in productivity is a promising task. Finally, stripping down the model to its essential ingredients is an important step in building a model that is both in line with empirical evidence and also amenable to a thorough welfare analysis. Such a model is necessary to assess the suitability of trade and welfare measures in empirical work. We leave this important challenge for future work.

References

[1] A. Alesina, S. Ardagna, G. Nicoletti, and F. Schiantarelli. Regulation And Investment. *Journal of the European Economic Association*, 3(4):791–825, 2005.

[2] M.J. Alvarez-Pelaez and C. Groth. Too little or too much R&D? *European Economic Review*, 49(2):437–456, 2005.

[3] L.G. Arnold. Growth, welfare, and trade in an integrated model of human-capital accumulation and research. *Journal of Macroeconomics*, 20(1):81–105, 1998.

[4] L.G. Arnold. The dynamics of the Jones R & D growth model. *Review of Economic Dynamics*, 9(1):143–152, 2006.

[5] L.G. Arnold. A generalized multi-country endogenous growth model. *International Economics and Economic Policy*, 4(1):61–100, 2007.

[6] B.Y. Aw, S. Chung, and M.J. Roberts. Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China). *The World Bank Economic Review*, 14(1):65–90, 2000.

[7] R. Baldwin. Regulatory protectionism, developing nations and a two-tier world trade system. 2000.

[8] R.E. Baldwin and F. Robert-Nicoud. Trade and growth with heterogeneous firms. *Journal of International Economics*, 74(1):21–34, 2008.

[9] R.J. Barro. Are Government Bonds Net Wealth? *Journal of Political Economy*, 82(6):1095, 1974.

[10] J.P. Benassy. Is there always too little research in endogenous growth with expanding product variety? *European Economic Review*, 42(1):61–69, 1998.

[11] A.B. Bernard and J. Bradford Jensen. Exceptional exporter performance: cause, effect, or both? *Journal of International Economics*, 47(1):1–25, 1999.

[12] A.B. Bernard, J. Eaton, J.B. Jensen, and S. Kortum. HPlants and Productivity in International Trade". *American Economic Review*, 93(4):1268–1290, 2003.

[13] A.B. Bernard, J.B. Jensen, and P.K. Schott. Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of US manufacturing plants. *Journal of International Economics*, 68(1):219–237, 2006.

[14] B. Büttner. Entry barriers and growth. *Economics Letters*, 93(1):150–155, 2006.

[15] L.M.B. Cabral and J. Mata. On the Evolution of the Firm Size Distribution: Facts and Theory. *The American Economic Review*, 93(4):1075–1090, 2003.

[16] S.K. Clerides, S. Lach, and J.R. Tybout. Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics*, 113(3):903–947, 1998.

[17] A.K. Dixit and J.E. Stiglitz. Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297–308, 1977.

[18] J. Eaton and S. Kortum. Trade in capital goods. *European Economic Review*, 45(7):1195–1235, 2001.

[19] C.J. Erceg, L. Guerrieri, and C. Gust. Trade adjustment and the composition of trade. *Journal of Economic Dynamics and Control*, 2007.

[20] G.M. Grossman and E. Helpman. *Innovation and Growth in the Global Economy*. MIT Press, 1991.

[21] P. Gustafsson and P. Segerstrom. Trade Liberalization and Productivity Growth. *CEPR Discussion Paper*, 5894, 2007.

[22] H. Hopenhayn. Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5):1127–1150, 1992.

[23] H. Hopenhayn. Exit, Selection, and the Value of Firms. *Journal of Economic Dynamics and Control*, 16(1):621–653, 1992.

[24] C.I. Jones. R & D-Based Models of Economic Growth. *Journal of Political Economy*, 103(4):759, 1995.

[25] W. Kornprobst. *Innovationsbasierte Wachstumstheorie*. Dissertation, University of Regensburg, Germany, 2008.

[26] C.W. Li. Endogenous vs. Semi-endogenous Growth in a Two-R&D-Sector Model. *The Economic Journal*, 110(462):109–122, 2000.

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

[28] G. Nicoletti and S. Scarpetta. Regulation, productivity and growth: OECD evidence. *Economic Policy*, 18(36):9–72, 2003.

[29] N. Pavcnik. Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants. *The Review of Economic Studies*, 69(1):245–276, 2002.

[30] L.A. Rivera-Batiz and P.M. Romer. Economic Integration and Endogenous Growth. *Quarterly Journal of Economics*, 106(2):531–555, 1991.

[31] M. Roberts and J.R. Tybout. The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs. *American Economic Review*, 87(4):545–564, 1997.

[32] P.M. Romer. Endogenous Technological Change. *Journal of Political Economy*, 98(S5):71, 1990.

[33] P.A. Samuelson. The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments. *Economic Journal*, 64(254):264–289, 1954.

[34] G. Sorger. Quality-improving horizontal innovations.

[35] P.J.G. Tang and K. Wälde. International competition, growth and welfare. *European Economic Review*, 45(8):1439–1459, 2001.

[36] J. Tirole. *The Theory of Industrial Organization*. MIT Press, 1988.

[37] J.R. Tybout. Plant-and Firm-Level Evidence on" New" Trade Theories. *Handbook of International Trade*, 2003.

[38] A. Young. Growth without Scale Effects. *Journal of Political Economy*, 106(1):41–63, 1998.

6 Appendix

6.1 Derivation of ν^*, z^*, l^*, and γ^*

Combining (69) and (71) yields

$$\begin{aligned}
\left(\frac{1}{\alpha}-\frac{1}{\sigma}-\frac{1}{\chi}-\frac{1}{\alpha}\right)n+\alpha\left(\frac{\alpha}{\sigma}-1\right)z^*+\alpha z^* &= \frac{\rho}{\sigma}\\
-\left(\frac{1}{\sigma}+\frac{1}{\chi}\right)n+\left[\alpha\left(\frac{\alpha}{\sigma}-1\right)+\alpha\right]z^* &= \frac{\rho}{\sigma}\\
\frac{\rho}{\sigma}+\left(\frac{1}{\sigma}+\frac{1}{\chi}\right)n &= z^*\frac{\alpha^2}{\sigma}\\
\rho+\left(1+\frac{\sigma}{\chi}\right) &= \alpha^2 z^*,
\end{aligned}$$

and hence

$$\begin{aligned}
z^* &= \frac{1}{\alpha^2}\left[\rho+n+\frac{\sigma}{\chi}n+\frac{(\sigma-1)n}{\chi}-\frac{(\sigma-1)n}{\chi}\right]\\
&= \frac{1}{\alpha^2}\left[\Delta+\frac{(\chi+\sigma)n-(\sigma-1)n}{\chi}\right]\\
&= \frac{1}{\alpha^2}\left[\Delta+\frac{1+\chi}{\chi}\right].
\end{aligned}$$

Using this expression in (69) yields ν^*.

$$\begin{aligned}
\frac{n}{\alpha}+\nu^* &= \frac{1}{\alpha}\left[\Delta+\frac{1-\chi}{\chi}n\right]\\
\nu^* &= \frac{1}{\alpha}\left[\Delta+\frac{1+\chi}{\chi}n-n\right]\\
\nu^* &= \frac{1}{\alpha}\left[\Delta+\frac{n}{\chi}\right]
\end{aligned}$$

This gives $\frac{\chi\tilde{l}}{\phi F_L}=n+\frac{\chi}{\alpha}\left[\Delta+\frac{n}{\chi}\right]$. Hence,

$$\begin{aligned}
\tilde{l}^* &= \frac{\phi F_L}{\chi}\left[n+\frac{\chi\Delta}{\alpha}+\frac{n}{\alpha}\right]\\
&= \phi F_L\left[\frac{n}{\chi}\left(1+\frac{1}{\alpha}\right)+\frac{\Delta}{\alpha}\right]\\
&= \frac{\phi F_L}{\alpha}\left[n\frac{1+\alpha}{\chi}+\Delta\right].
\end{aligned}$$

Finally, from (65),

$$\left[\frac{\rho}{\sigma}-\left(1-\frac{1}{\sigma}n\right)\right]-\left(\frac{\alpha^2}{\sigma}-1\right)\frac{1}{\alpha^2}\left(\frac{1+\chi}{\chi}n+\Delta\right)=\gamma^*.$$

After rearranging and canceling terms, we have

$$\begin{aligned}
\gamma^* &= \frac{\rho}{\sigma}-n+\frac{n}{\sigma}-\left(\frac{1}{\sigma}-\alpha^2\right)\left(\frac{1+\chi}{\chi}n+\frac{\sigma-1}{\chi}n+\rho\right)\\
&= \frac{\rho}{\sigma}-n+\frac{n}{\sigma}-\frac{1+\chi}{\chi}\frac{n}{\sigma}-\frac{\sigma-1}{\chi}\frac{n}{\sigma}-\frac{\rho}{\sigma}+\frac{1+\chi}{\chi}\alpha^2 n+\frac{\sigma-1}{\chi}\alpha^2 n+\alpha^2\rho\\
&= \frac{1}{\alpha^2}\left[\alpha^2\left(\frac{\rho}{\sigma}-n+\frac{n}{\sigma}-\frac{1+\chi}{\chi}\frac{n}{\sigma}-\frac{\sigma-1}{\chi}\frac{n}{\chi}-\frac{\rho}{\sigma}\right)+\left(\frac{1+\chi}{\chi}+\frac{\sigma-1}{\chi}\right)n+\rho\right]\\
&= \frac{1}{\alpha^2}\left[\alpha^2 n\left(\frac{1}{\sigma}-1-\frac{1+\chi}{\chi}\frac{1}{\sigma}-\frac{\sigma-1}{\chi\sigma}\right)+\frac{\chi+\sigma}{\chi}n+\rho\right]\\
&= \frac{1}{\alpha^2}\left[\alpha^2 n\left(\frac{\chi-\chi\sigma-(1+\chi)-\sigma+1}{\chi\sigma}\right)+\frac{\chi+\sigma}{\chi}n+\rho\right]\\
&= \frac{1}{\alpha^2}\left[-\alpha^2 n\left(\frac{1+\chi}{\chi}\right)+\frac{\chi+\sigma}{\chi}n+\rho\right]\\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\left[-\alpha^2(1+\chi)+\chi+\sigma\right]+\rho\right\}\\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\left[(1-\alpha^2)\chi-\alpha^2+\sigma\right]+\rho\right\}\\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\sigma+n\left[(1-\alpha^2)-\frac{\alpha^2}{\chi}\right]+\rho\right\}\\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\sigma-\frac{n}{\chi}+\rho+\frac{n}{\chi}+n\left[(1-\alpha^2)-\frac{\alpha^2}{\chi}\right]\right\}\\
&= \frac{1}{\alpha^2}\left\{\Delta+\frac{n}{\chi}\left[1+\chi(1-\alpha^2)-\alpha^2\right]\right\}\\
&= \frac{1}{\alpha^2}\left[\Delta+(1-\alpha^2)n\frac{1+\chi}{\chi}\right].
\end{aligned}$$