Chapter 3
Product Variety

April 2, 2008

1 Introduction

The inability of the AK paradigm to produce a convincing model of long-run growth and convergence motivated a second wave of endogenous growth theory, consisting of innovation-based growth models, which themselves belong to two parallel branches. One branch is the product-variety model of Romer (1990), according to which innovation causes productivity growth by creating new, but not necessarily improved, varieties of products.

The other branch of innovation-based theory, developed by Aghion and Howitt (1992), grew out of modern industrial organization theory, and is commonly referred to as "Schumpeterian" growth theory, because it focuses on quality-improving innovations that render old products obsolete, and hence involves the force that Schumpeter called "creative destruction."

This chapter presents Romer's product-variety model. Subsequent chapters will develop the Schumpeterian version of endogenous growth theory. Section 2 below presents a simple product-variety model in which final output is used as R&D input. Section 3 presents a variant in which labor is used as R&D input. Section 4 confronts the model with empirical evidence.

2 Endogenizing technological change

The model we present in this section builds on the idea that productivity growth comes from an expanding variety of specialized intermediate products. Product variety expands gradually because discovering how to produce a larger range of products takes real resources, including time. The model formalizes an old idea that goes back to Young (1928), namely that growth is induced and sustained by increased specialization.

For each new product there is a sunk cost of product innovation that must be incurred just once, when the product is first introduced, and never again. The sunk costs can be thought of as costs of research, an activity that results in innovations that add to the stock of technological knowledge. In this case, technological knowledge consists of a list of blueprints, each one describing how to produce a different product, and every innovation adds one more blueprint to the list.

What makes this different from an AK model is not just the sunk cost of product development, but also the fact that fixed costs make product markets monopolistically competitive rather than perfectly competitive. Imperfect competition creates positive profits, and these profits act as a reward to the creation of new products. This is important because it allows the economy to overcome the problem created by Euler's theorem, which we discussed in the previous chapter; that is, the problem that under perfect competition all of output would go to those who supplied K and L, with nothing left over to compensate those who provide the technological knowledge underlying A.

2.1 A Simple variant of the product-variety model

There is a fixed number L of people, each of whom lives forever and has a constant flow of one unit of labor that can be used in manufacturing. For simplicity we suppose that no one has a demand for leisure time, so each person offers her one unit of labor for sale inelastically (that is, no matter what the wage rate). Her utility each period depends only on consumption, according to the same isoelastic function that we presented in connection with the Cass-Koopmans-Ramsey model in Chapter 1:

$$u(c) = \frac{c^{1-\varepsilon}}{1-\varepsilon},\ \varepsilon > 0$$

and she discounts utility using a constant rate of time preference ρ. As we saw in Chapter 1, this means that in steady state the growth rate g and the interest rate r must obey the Euler equation, which can be written as:

$$g = \frac{r-\rho}{\varepsilon} \tag{1}$$

Final output is produced under perfect competition, using labor and a range of intermediate inputs, indexed by i in the interval $[0, M_t]$, where M_t is our measure of product variety. The final-good production function at each date t is

$$Y_t = L^{1-\alpha} \int_0^{M_t} x_i^{\alpha} di \tag{2}$$

where Y_t is output, and each x_i is the amount of intermediate product i used as input. Labor input is always equal to the fixed supply L. The coefficient α lies between zero and one.[1]

Each intermediate product is produced using the final good as input, one-for-one. That is, each unit of intermediate product i produced requires the input of one unit of final good.

According to (2), product variety enhances overall productivity in the economy. To see how this works, let X_t be the total amount of final good used in producing intermediate products. According the the one-for-one technology, X_t must equal total intermediate output:

$$X_t = \int_0^{M_t} x_i di$$

Now suppose that each intermediate product is produced in the same amount x. (This will indeed be the case in equilibrium, as we will see shortly.) Then $x = X_t/M_t$. Substituting this into the production function (2) yields:

$$Y_t = L^{1-\alpha} \int_0^{M_t} (X_t/M_t)^{\alpha} \, di = M_t^{1-\alpha} L^{1-\alpha} X_t^{\alpha} \tag{3}$$

which is increasing in M_t given the factor inputs L and X_t:

$$\partial Y_t / \partial M_t = (1-\alpha) \, Y_t/M_t > 0$$

The final good is used for consumption and investment (in producing blueprints). Its only other use is in producing intermediate products. So the economy's Gross Domestic Product (GDP) is final output Y_t minus the amount used in intermediate production:

$$GDP_t = Y_t - X_t \tag{4}$$

Each intermediate product is monopolized by the person that created it. The monopolist seeks to maximize the flow of profit at each date, measured in units of final good:

$$\Pi_i = p_i x_i - x_i$$

where p_i is the price in units of final good. That is, her revenue is price times quantity and her cost is equal to her output, given the one-for-one technology.

Since the price of an input to a perfectly competitive industry is the value of its marginal

[1] We attach a time subscript t to M_t and Y_t because product variety and final-good production will be growing over time in a steady state. However, we leave the time subscript off x_i because, as we will see, the output of each intermediate product will be constant over time.

product, therefore we have:[2]

$$p_i = \partial Y_t / \partial x_i = \alpha L^{1-\alpha} x_i^{\alpha-1} \tag{5}$$

Therefore the monopolist's profit depend on her output according to:

$$\Pi_i = \alpha L^{1-\alpha} x_i^{\alpha} - x_i$$

She will choose x_i so as to maximize this expression, which implies the first-order condition

$$\partial \Pi_i / \partial x_i = \alpha^2 L^{1-\alpha} x_i^{\alpha-1} - 1 = 0$$

It follows that the equilibrium quantity will be the same constant in every sector i:

$$x = L\alpha^{\frac{2}{1-\alpha}} \tag{6}$$

and so will the equilibrium profit flow:[3]

$$\Pi = \frac{1-\alpha}{\alpha} L\alpha^{\frac{2}{1-\alpha}}. \tag{7}$$

Substituting $X_t = M_t x$ into the production relation (3), and then again into the definition (4), we see that final-good output and the economy's GDP will both be proportional to the degree of product variety:

$$Y_t = M_t L^{1-\alpha} x^{\alpha}$$

$$GDP_t = M_t \left(L^{1-\alpha} x^{\alpha} - x \right)$$

[2] Strictly speaking, the derivative $\partial Y_t / \partial x_i$ makes no mathematical sense, because a change in a single x_i would have no measurable effect on the integral in (2). But it does make economic sense, because our assumption that there is a continuum of intermediate products is itself just an approximation. What we are really saying is that output depends on the sum of contributions from a large discrete number M of intermediate products, each of which makes a small contribution:

$$Y_t = L^{1-\alpha} \Sigma_1^M x_i^{\alpha} \tag{Y}$$

and we are approximating this production function by assuming a continuum of products. The derivative of the production function (Y) is indeed given by expression (5).

[3] By definition

$$\Pi_i = (p_i - 1)\, x_i$$

Using (5) to substitute for p_i and then (6) to substitute for x_i we get

$$\Pi_i = \left[\alpha L^{1-\alpha} \left(L\alpha^{\frac{2}{1-\alpha}} \right)^{\alpha-1} - 1 \right] L\alpha^{\frac{2}{1-\alpha}}$$

Expression (7) follows directly from this because the term in square brackets equals $\frac{1-\alpha}{\alpha}$.

Therefore the growth rate of GDP will be the proportional growth rate of product variety:

$$g = \frac{1}{M_t} \frac{dM_t}{dt}$$

Product variety grows at a rate that depends on the amount R_i of final output that is used in research. That is, the output of research each period is the flow of new blueprints, each of which allows a new product to be developed. So we have

$$dM_t/dt = \lambda R_t$$

where λ is a (positive) parameter indicating the productivity of the research sector.

Assume that the research sector of the economy is perfectly competitive, with free entry. Then the flow of profit in the research sector must be zero. Each blueprint is worth Π/r to its inventor, which is the present value of the profit flow Π discounted at the market interest rate r. Hence the flow of profit in research is

$$(\Pi/r)\,\lambda R_t - R_i$$

which is just the flow of revenue (output λR_t time price Π/r) minus cost R_i. For this to be zero we need a rate of interest that satisfies the "research-arbitrage equation:"

$$r = \lambda \Pi$$

That is, the rate of interest must equal the flow of profit that an entrepreneur can receive per unit invested in research.

Substituting from the research-arbitrage equation into the Euler equation (1) we have

$$g = \frac{\lambda \Pi - \rho}{\varepsilon}$$

Substituting the expression (7) for Π in this equation yields the following expression for the equilibrium growth rate as a function of the primitive parameters of the model:

$$g = \frac{\lambda \frac{1-\alpha}{\alpha} L \alpha^{\frac{2}{1-\alpha}} - \rho}{\varepsilon}.$$

We immediately see that growth increases with the productivity of research as measured by the parameter λ and with the size of the economy as measured by labor supply L, and decreases with the rate of time preference ρ.

The prediction that g should increase with L was first seen as a virtue of the model, suggesting that larger countries or larger free trade zones should grow faster. However, Jones (1995) pointed out that this prediction is counterfactual, to the extent that the number of researchers has substantially increased in the US over the period since 1950, whereas the growth rate has remained on average at 2% over the same period. We shall come back to this debate on "scale effects" in more detail in Chapter 4 below.

2.2 The Romer model with labor as R&D input

The original Romer model supposed that labor was the only R&D input.[4] To see how the model works under this alternative assumption, we now suppose that labor can be used either in manufacturing the final good (L_1) or alternatively in research (L_2). Labor used in these two activities must add up to the total labor supply L, which we again assume to be a given constant. So:

$$L = L_1 + L_2.$$

We restrict attention to steady states in which L_1 and L_2 are both constant.

Final output is produced by labor and intermediates according to the same production function as before:

$$Y_t = L_1^{1-\alpha} \int_0^{M_t} x_i^{\alpha} di, \tag{8}$$

where the labor input is now L_1 instead of the total labor supply L. The price of each intermediate product is again its marginal product in the final sector:

$$p = \alpha L_1^{1-\alpha} x^{\alpha-1}.$$

Each intermediate product is again produced one-for-one with final output, so the profit flow to each intermediate monopolist is given by:

$$\begin{aligned} \Pi &= \max_x \{px - x\} \\ &= \max_x \left\{\alpha L_1^{1-\alpha} x^{\alpha} - x\right\} \end{aligned}$$

[4] Actually, Romer interpreted the R&D input as "human capital", but since he took its total supply as given this difference is purely terminological. Romer also supposed that intermediate products were produced by capital, which for simplicity we ignore in this presentation. We defer the integration of endogenous technical change and capital accumulation until chapter 5 below, and note here that nothing of importance is lost by ignoring capital accumulation in the Romer model.

which implies the profit-maximizing quantity:

$$x = L_1 \alpha^{\frac{2}{1-\alpha}}. \tag{9}$$

and the equilibrium profit flow:[5]

$$\Pi = \frac{1-\alpha}{\alpha} L_1 \alpha^{\frac{2}{1-\alpha}}. \tag{10}$$

The measure of product variety M_t now grows at a rate that depends upon the amount L_2 of labor devoted to research, according to:

$$dM_t/dt = \lambda M_t L_2.$$

This equation reflects the existence of spillovers in research activities: that is, all researchers can make use of the accumulated knowledge M_t embodied in existing designs. Note that there are *two* major sources of increasing returns in this model: specialization or product differentiation and research spillovers.

The flow of profit in research is now

$$(\Pi/r)\, \lambda M_t L_2 - w_t L_2$$

where w_t is the equilibrium wage rate that must be paid to researchers. Setting this flow equal to zero yields the research-arbitrage equation for this version of the model:

$$r = \lambda M_t \Pi / w_t$$

which again states that the rate of interest must equal the flow of profit that the entrepreneur can receive from investing one unit of final good into research; i.e. from using the services of $(1/w_t)$ units of research labor and thereby producing $\lambda M_t\,(1/w_t)$ blueprints each worth Π per period.

To make use of this research-arbitrage equation we need to solve for the equilibrium wage rate w_t. Since the final sector is perfectly competitive, w_t equals the marginal product of labor, which can be calculated as follows. Since each intermediate sector produces the same

[5] Equations (9) and (10) are the same as equations (6) and (7) above, except with L_1 in the place of L. The former can be derived from the first-order condition of profit maximization using the same logic as in the previous section, and the latter can be derived using the logic of footnote 3 above.

constant output x, the production function (8) implies that:

$$Y_t = L_1^{1-\alpha} M_t x^{\alpha}. \tag{11}$$

Therefore:

$$w_t = \frac{\partial Y_t}{\partial L_1} = (1-\alpha) L_1^{-\alpha} M_t x^{\alpha}$$

which can be written, using expression (9), as

$$w_t = (1-\alpha) \alpha^{\frac{2\alpha}{1-\alpha}} M_t \tag{12}$$

So the research-arbitrage equation can be rewritten, using (10) to substitute for Π and using (12) to substitute for w_t, as:

$$r = \alpha \lambda L_1$$

Since

$$g = \frac{1}{M_t} \frac{dM_t}{dt} = \lambda L_2 = \lambda (L - L_1)$$

we have

$$r = \alpha (\lambda L - g)$$

Substituting this expression for r into the Euler equation (1) yields

$$g = \frac{\alpha \lambda L - \rho}{\alpha + \varepsilon}.$$

So again, growth increases with the productivity of research activities λ and with the size of the economy as measured by total labor supply L, and decreases with the rate of time preference ρ. Furthermore, both because intermediate firms do not internalize their contribution to the division of labor (i.e. to product diversity) and because researchers do not internalize research spillovers, the above equilibrium growth rate is always *less* than the social optimum.[6]

[6] Benassy (1998) shows, however, that with a slightly more general form of the product-variety model, the equilibrium growth rate could exceed the optimal rate.

3 From theory to evidence

3.1 Estimating the effect of variety on productivity

A recent paper by Broda, Greenfield and Weinstein (2006), henceforth BGW, exploits trade data to test for the effects of product variety on productivity levels and growth. If we believe in the mechanism described above, trade should raise productivity because producers gain access to new imported varieties of inputs (the level effect); moreover, trade and the resulting increase in input variety should reduce the cost of innovation and thus result in more variety creation in the future (the growth effect). Also, the impact of increased product variety on productivity should depend upon the elasticity of substitution among different varieties of a good, and/or upon shifts in expenditures shares among new, remaining and disappearing goods. In particular, increasing the number of varieties should not have much of an effect on productivity if new varieties are close substitutes to existing varieties or if the share of new varieties is small relative to existing ones.

BGW analyze bilateral trade flows between 73 countries over the period 1994-2003. Using import data in the 6-digit Harmonized System (HS) product categories, they compute elasticities of substitution and supply for about 200 import sectors in each country. They consider the production function

$$Y_t = (A_t L_t)^{1-\alpha} \left(\sum_1^{M_t} x_{it}^{\upsilon}\right)^{\alpha/\upsilon},$$

where $\alpha \in (0,1)$ is one minus the share of labor in output and $\upsilon \in (0,1)$ measures the elasticity of substitution between varieties of input goods x_{it}, with a higher υ corresponding to more substitutable inputs.

If we focus on equilibria where all input goods are used with the same intensity x_t, the above equation becomes:

$$Y_t = (A_t L_t)^{1-\alpha} M_t^{\alpha/\upsilon} x_t^{\alpha}.$$

If each input is produced one-for-one with capital, then

$$Y_t = (A_t L_t)^{1-\alpha} M_t^{(1-\upsilon)\alpha/\upsilon} K_t^{\alpha},$$

where $K_t = M_t x_t$ is the aggregate capital stock. Taking logs, we obtain:

$$\ln Y_t = (1-\alpha)(\ln A_t + \ln L_t) + \alpha \ln K_t + (1-\upsilon)\alpha/\upsilon \ln M_t.$$

Differentiating both sides with respect to time, we obtain:

$$\frac{\dot{Y}_t}{Y_t} = (1-\alpha)\frac{\dot{L}_t}{L_t} + \alpha\frac{\dot{K}_t}{K_t} + \widehat{B}_t,$$

where

$$\widehat{B}_t = (1-\alpha)\frac{\dot{A}_t}{A_t} + (1-\upsilon)\alpha/\upsilon\frac{\dot{M}_t}{M_t}$$

is total factor productivity (TFP) growth, also known as the Solow residual.[7] This measure of TFP growth has two components: a product-variety component captured by the term in $\dot{M}_t/M_t$, and a quality component embodied in the term in $\dot{A}_t/A_t$. BGW are primarily interested in the contribution of variety to total productivity growth.

According to the above equation, a lower υ (that, is a lower degree of substitutability between inputs) or a higher share of the intermediate goods α should result in a higher impact of increased variety on TFP.

BGW estimate elasticities separately for each good and importing country and then regress per capita GDP on these elasticities. They find no strong relationship between income per capita and the elasticity of substitution across countries. The typical (median) country experienced a net increase in varieties of 7.1 percent over 10 years in the typical (median) sector, which is about 0.7 percent per year. BGW then show that the growth in new varieties over the period 1994-2003 increased productivity by 0.13 percent per year for the typical country in the sample.

The relationship between variety and productivity is even lower for developed countries: "most of the productivity growth in many of the largest countries cannot be accounted for by new imports" (BGW p. 21). In particular the US have the second smallest gain among developed countries from imported variety. BGW summarize their findings as follows: "The median developed country's productivity growth was about 2 percent per year, but the median contribution of imported variety growth to productivity was only 0.1 percent per year, suggesting that for the typical developed country, new imported varieties are only a small part of the story behind their productivity growth. The impact of new varieties on developing countries is substantially higher. The typical developing country saw its productivity rise by 0.13 to 0.17 percent per year (depending on the sample) due to new imported varieties" (BGW p.22).

One might argue that BGW underestimate the effects of variety in more developed countries. In particular their approach does not take into account inputs that are not imported, and these are likely to be more numerous in developed countries. Nevertheless, the effects

[7] TFP and the Solow residual will be explained in more detail in Chapter 5.

of variety on productivity and on productivity growth appear to be relatively small, even in less developed economies.

3.2 The importance of exit in the growth process

An important limitation of the product-variety model is that it assumes away obsolescence of old intermediate inputs. Indeed, if old intermediate inputs were to disappear over time, the variety term in the above Solow residual would go down, and thus so would the economy's per capita GDP.

In ongoing work with Pol Antras and Susanne Prantl, we have combined UK establishment-level panel data with the input-output table to estimate the effect on TFP growth arising from growth in high-quality input in upstream industries, and also from exit of obsolete input-producing firms in upstream industries. Specifically, we take a panel of 23,886 annual observations on more than 5,000 plants in 180 4-digit industries between 1987 and 1993, together with the 1984 UK input-output table, to estimate an equation of the form:

$$g_{ijt} = \beta_0 + \beta_1 entry_{jt-1} + \beta_2 exit_{jt-1} + \beta_3 Z_{ijt-1} + est_i + ind_j + yr_t + u_{ijt}$$

where g_{ijt} is the productivity growth rate of firm i in industry j. The first regressor $entry_{jt-1}$ is the entry measure, calculated as the increase in the fraction of input to the production of good j which is provided by foreign firms (foreign firms are more likely to account for entry that takes place at frontier technological level). The second regressor $exit_{jt-1}$ is our measure of exit of obsolete upstream input-producing firms: we use the fraction of employment accounted for by upstream exiting firms, thereby putting more weight on large exiting firms than on small ones. Establishment (est_i), industry (ind_j) and year (yr_t) effects are included, along with the other controls in Z_{ijt-1}, including a measure of the plant's market share.

The result of this estimation is a significant positive effect of both upstream quality improvement and upstream input-production exit. These results are robust to taking potential endogeneity into account by applying an instrumental variable approach, using instruments similar to those of Aghion et al (2005). The effects are particularly strong for plants that use more intermediate inputs; i.e., plants with a share of intermediate product use above the sample median. Altogether, the results we find are consistent with the view that quality-improving innovation is an important source of growth. The results are however not consistent with the horizontal innovation model, in which there should be nothing special about the entry of foreign firms, and according to which the exit of upstream firms should if anything reduce growth by reducing the variety of inputs being used in the industry.

Comin and Mulani (2005) have produced additional evidence to the effect that exit as well as entry is important to the growth process. Using a sample of US firms they show that, according to two measures of turnover in industry leadership that they construct, turnover is positively related to earlier R&D. Again, this is evidence of a creative-destruction element to the innovation process that one would not expect to find if the primary channel through which innovation affected economic growth was by increasing product variety. Indeed the product-variety theory has little to say at all about how productivity varies across firms in an industry, let alone how the productivity ranking would change over time.

In addition to these results, Fogel, Morck and Yeung (2005) have produced evidence to the effect that innovation is linked to the turnover of dominant firms. Using data on large corporate sectors in 44 different countries over the 1975-96 period, they find that economies whose top 1975 corporations declined more grow faster than other countries with the same initial per-capita GDP, level of education and capital stock. Again, this evidence of an association between growth and enterprise turnover has no counterpart in the horizontal-innovation theory.

In order to formalize the notion of (technical or product) obsolescence, one needs to move away from *horizontal* models of product development à la Dixit and Stiglitz (1977) into *vertical* models of quality improvements, which we do in the next chapter.

4 Literature Notes

Romer (1987) developed a precursory model of growth with expanding variety, where growth is sustained in the long run by the fact that output is produced with an expanding set of inputs, which in turn prevents aggregate capital to run into decreasing returns. The improved division of labor over time, which itself is made possible by output growth, which makes it possible to pay the fixed costs of producing an ever expanding set of inputs. Romer used the framework of monopolistic competition introduced by Dixit and Stiglitz (1977) and extended by Ethier (1982). Romer (1990) completes the description of the product-variety model by introducing an R&D sector which generates blueprints for new inputs as a result of voluntary profit-motivated horizontal innovations. Technological change was thereby endogenized.

The framework has been extended in several directions, and we refer the reader to the excellent Handbook chapter by Gancia and Zilibotti (2005). Grossman and Helpman (1991, Chapter 3) present a didactic treatment of a framework with expansion of consumer products that enter the utility function (as in Spence, 1976) instead of intermediate products entering the production function. Rivera-Batiz and Romer (1991) and Grossman and Helpman (1991) have used the framework to analyze the effects of market integration on growth. More recently, the idea of directed technological change was integrated by Acemoglu and Zilibotti

(2001) into a framework of growth with expanding variety, to analyze implications of the skill-technology complementarity for the persistence of productivity differences across countries.

Jones (1995) challenges the model for generating scale effects for which he could not any evidence based on US time series (see Chapter 4 below). On the other hand, Kremer (1993) argued the hypothesis of a positive relation between world per capita growth and world population (or the world aggregate output) might be correct over long, very long periods of time.