INDEX

Preface 2

In memoriam Theo Leers and Kees Gorter 4

Programme NAKE workshop 10-14 December 2001
Wageningen 5

Workshop report:
"Pricing"
by Remco van der Molen 8

Workshop report:
"Computational Economics"
by Laurens Swinkels 18

Workshop report:
"Innovation-based growth theory"
by Egbert Jongen 32

Upcoming 45

Preface

The end of yet another year is approaching very fast. For NAKE, it has been a good year, with many interesting courses, a great workshop in June, and a successful research day in October.

But the year is not over yet. We still have one workshop to go this month in Wageningen As always, four distinguished economists will lecture at this workshop. However, this time there is a small, but not unimportant change in the tradition: for the first time in 15 years a female economist will teach a course at a NAKE workshop. Of course, in itself this is not an important fact. There is quite a number of female economists and there is no reason why they should not teach at a workshop. But the fact that it took 15 years before it actually happened makes it quite special. **Susanne Lohmann** will be the first. I hope many will follow (and not only once every fifteen years). The other lecturers at this month's workshop are **Russel Cooper**, **Christian Gollier,** and **Paul Ruud**. You will find more information on the programme of the workshop from page 5.

In this *NAKE Nieuws* we commemorate two appreciated colleagues who recently passed away: **Theo Leers** (KUB) and **Kees Gorter** (VU). You will find an *In Memoriam* for Theo and for Kees on page 4.

This issue also contains three reports on the workshop in June: **Laurens Swinkels** (KUB) reports on Ken Judd's inspiring lectures on numerical methods, **Remco van der Molen** (RUG) writes on "Pricing" based on Preston McAfee's course, and **Egbert Jongen** (VU) summarises Peter Howitt's overview of innovation-based growth theory.

Finally, I would like to redress an omission in the previous issue of *NAKE Nieuws*. When mentioning the students who received the NAKE diploma at the workshop in June, I overlooked the name of **Michiel de Nooij**. Sorry Michiel!

Wishing you all a Merry Christmas and a wonderful 2002,

Lex Meijdam

PS. Note that we have added a letter to this *NAKE Nieuws* to update our mailing list. Please do not forget to return this letter. If you do not react we will assume that you are no longer interested in receiving *NAKE Nieuws*!

In Memoriam Theo Leers

On September 4, former NAKE-student Theo Leers suddenly passed away at the age of 33. Theo started as a PhD-student 1996. He was an enthusiastic NAKE-student and many of his colleagues cherish good remembrances of the evenings they spent with Theo at various workshops. Only a few months ago, in June 2001, Theo received the NAKE diploma. In that month he also defended his PhD-thesis *Public Pensions and Population Ageing; An Economic Analysis of Fertility, Migration, and Social-Security Policy* at Tilburg University. After being a PhD-student, Theo started as an assistant professor at this university. This seemed to be the start of a successful career in academia. One of the chapters of his thesis was already accepted for publication in the *Journal of Public Economics*. Unfortunately, it turned out differently. It is hard to accept the loss of such a fine colleague. Our thoughts are with his wife Elvera, his daughter Bélèn and his parents.

In Memoriam Kees Gorter

On October 26, NAKE-fellow Kees Gorter passed away at the age of 40. Kees was associate professor in the Department of Regional Sciences at the Vrije Universiteit Amsterdam. He was an enthusiastic researcher with a long list of publications in the field of regional science and labour economics. Furthermore, Kees was an animated teacher and a rich source of information for young researchers. In May, at Kees' 40th birthday, Arno van der Vlist, one of the PhD-students he advised, defended his thesis. Above all, Kees was a perfect colleague. Even during his spell of sickness, he kept close contact and visited the department as often as possible. All who knew Kees will lack his good sense of humor and inspiring optimism. Our thoughts are with his wife Ineke and his sons Maarten and Floris.

NAKE Workshop

10 - 14 December 2001

Wageningen University

During the week from Monday 10 December to Friday 14 December 2001 the Netherlands Network of Economics (NAKE) will organize its 31st PhD. workshop. Four distinguished economists teach intensive courses on microeconomics, macroeconomics, econometrics and public economics. Each course consists of five lectures spread out over five days.

Courses

Russel Cooper
Boston University
“Dynamic Programming”

Christian Gollier
Université de Toulouse
“The Economics of Risk and Time”

Susanne Lohmann
University of California, Los Angeles
“Why Some Groups Work and Others Don't”

Paul Ruud
University of California, Berkeley
“Limited Dependent Variable Models: Estimation with Simulation”

PROGRAMME NAKE WORKSHOP
WAGENINGEN, 10-14 DECEMBER 2001

Day	Time	Programme	Location
Monday December 10	10.00 – 11.00	*Registration / Coffee*	*Terras zaal*
	11.00 – 12.30	**Lohmann**	Roghorst zaal
	12.30 – 13.30	*Lunch*	*Terras zaal*
	13.30 – 14.45	**Lohmann**	Roghorst zaal
	14.45 – 15.00	*Break*	*Terras zaal*
	15.00 – 16.15	**Cooper**	Roghorst zaal
	16.15 – 16.30	*Break*	*Terraszaal*
	16.30 – 17.45	**Ruud**	Roghorst zaal
	17.45 – 19.15	*Welcome reception*	*Terraszaal*
Tuesday December 11	09.00 – 10.45	**Lohmann**	Roghorst zaal
	10.45 – 11.00	*Break*	*Terraszaal*
	11.00 – 12.45	**Ruud**	Roghorst zaal
	12.45 – 13.45	*Lunch*	*Terraszaal*
	13.45 – 15.30	**Cooper**	Roghorst zaal
	15.30 – 15.45	*Break*	*Terraszaal*
	15.45 – 17.30	**Lohmann**	Roghorst zaal
	19.30 - 20.00	Excursion: a walk and a drink in Wageningen	Bowlespark 1A

………

Day	Time	Programme	Location
Wednesday December 12	09.00 – 10.30	**Cooper**	Roghorst zaal
	10.30 – 10.45	*Break*	*Terras zaal*
	10.45 – 12.15	**Lohmann**	Roghorst zaal
	12.15 – 13.45	*Lunch*	*Café "De Tijd",* *Markt 12*
	13.45 – 15.15	**Gollier**	Roghorst zaal
	15.15 – 15.30	*Break*	*Terras zaal*
	15.30– 17.00	**Ruud**	Roghorst zaal
	17.00 – 18.30	Individual consultations	to be announced
Thursday December 13	09.00 – 10.45	**Gollier**	Roghorst zaal
	10.45 – 11.00	*Break*	*Terras zaal*
	11.00 – 12.45	**Cooper**	Roghorst zaal
	12.45 – 13.45	*Lunch*	*Terras zaal*
	13.45 – 15.30	**Ruud**	Roghorst zaal
	15.30 – 15.45	*Break*	*Self service restaurant*
	15.45 – 17.30	**Gollier**	*Roghorst zaal*
	20.00 -	*Workshop dinner*	*Indian Restaurant* *"The Rose"* *Hoogstraat 7*
Friday December 14	09.15 – 10.45	**Gollier**	Roghorst zaal
	10.45 – 11.00	*Break*	*Self service restaurant*
	11.00 – 12.30	**Ruud**	Roghorst zaal
	12.30 – 13.30	*Lunch*	*Terras zaal*
	13.30 – 15.00	**Cooper**	Roghorst zaal
	15.00 – 15.15	*Break*	*Self service restaurant*
	15.15 – 16.30	**Gollier**	Roghorst zaal
	16.30 - …	*Closing drinks*	To be announced

Pricing
R. Preston McAfee

Report by Remco van der Molen*

1 Introduction

The lectures of professor McAfee introduced us to a range of topics in pricing. He started his series of lectures with some highly interesting examples, indicating the practical relevance of the topics. On the one hand, the theories that professor McAfee introduced us to explained why it may be perfectly rational for a manufacturer to deliberately produce a version of his product that is clearly inferior to the original, and sell it at a lower price. On the other hand, the theories also proved very useful in many (industrial) policy issues related to the (supposed) abuse of monopoly power. In the following, I give a brief review of the most important issues that were dealt with. These issues are price discrimination, quantity discounts, quality premia, peak-load pricing, priority pricing, price dispersion, and durable good pricing.

2 Price discrimination

Examples abound in which the same economic good is sold at different prices to different consumers. In general, this phenomenon, known as price discrimination, can be seen as an attempt by a monopolist to appropriate a larger part of the economic surplus.

Assume there is a single good, with downward sloping demand $p(q)$, and a profit maximizing monopolist. Costs are normalized to zero. In case of a uniform price, the monopolist earns $q_0 p(q_0)$. A two price discriminating monopolist earns $q_1 p(q_1) + (q_2 - q_1) p(q_2)$, where the q's are the optimal quantities. Note that, in this case, welfare can only be increased by selling a higher quantity, as this reduces the inefficiency due to the

---

*Department of Economics, University of Groningen, PO Box 800, 9700 AV Groningen, r.m.van.der.molen@eco.rug.nl

monopolist's quantity rationing. It can easily be shown (see Varian, 1985) that welfare is higher under price discrimination. The intuition is as follows. In the non-discriminatory case, in determining the optimal quantity the monopolist maximizes over all consumers' willingnesses to pay. When discriminating, the monopolist sells q_1 to the consumers with a high willingness to pay. In determining how much to sell to the low-valuation consumers, the monopolist faces a new optimization problem, where the highest valuation consumers can be ignored. This will lead to a total quantity sold of $q_2 > q_0$, thus increasing welfare.

This is not a general result, however. If a monopolist faces n markets, with demand in market i given by $x_i(\mathsf{p})$, where p is the price vector, price discrimination may or may not increase welfare. Note that, in this case, demand is interdependent between different markets. Moreover, it is assumed that the cross-price elasticities are symmetric, i.e. $dx_i/dp_j = dx_j/dp_i$. Simple optimization leads to a generalization of the inverse elasticity rule: $L = -E^{-1}\mathsf{1}$, where L is the n-vector of Lerner indices $(p_i - mc)/p_i$, and E contains the own price elasticities (on the diagonal) and the cross-price elasticities (off the diagonal). So, the price/cost margins are determined by the inverse of the matrix of elasticities.

Charging different prices p_i has two effects. The first effect is that output is reallocated across markets. This causes the marginal rate of substitution to differ among consumers. Since a given quantity is optimally distributed by charging a uniform price, this has a negative impact on welfare. The second effect is a change in output; a higher output reduces the monopoly pricing distortion, thereby increasing total welfare. Thus, price discrimination may or may not lead to higher output. For example, output may be increased by serving markets that were not served under uniform pricing because the price was too high. The two effects imply that a necessary condition for price discrimination to increase welfare is that it increases total output.

2.1 Ramsey pricing

Consider a social planner whose objective is to maximize total welfare subject to a break-even constraint for the monopolist. Which prices would be set by the social planner and how does this differ from the prices that would be set by a monopolist? The planner faces the following problem:

$$\max_{p_i} u(\mathsf{x}) - c(\mathsf{x}) \text{ s.t. } \mathsf{px} - c(\mathsf{x}) \geq \pi_0$$

Writing the Lagrangian and solving for the first order condition gives the Ramsey price solution

$$-\frac{\lambda}{1+\lambda} = \sum_{j=1} \frac{p_j - mc}{p_j}\epsilon_{ij},$$

where λ, the Lagrange multiplier is the marginal increase in welfare associated with a decrease in firm profits, and ϵ_{ij} is the cross-price elasticity of substitution

$$\epsilon_{ij} = \frac{p_j}{x_i}\frac{dx_i}{dp_j}$$

Note the similarity with the optimizing monopolist outcome. The social planner would also set prices according to some inverse elasticity rule. So, whether or not the monopoly prices will be implemented depends on λ. If $\lambda \to \infty$, the Ramsey price converges to the monopoly outcome, whereas $\lambda = 0$ gives a markup of 0, i.e. marginal cost pricing. The Ramsey solution is thus in between marginal cost pricing and monopoly pricing, depending on λ.

2.2 Arbitrage

The interdependencies of demand depend crucially on the possibility of arbitrage. In a situation where there is no possibility whatsoever for arbitrage, the cross-price elasticity will be zero at all prices. This means that the matrix of elasticities E becomes diagonal. In this case, the inverse elasticity rule reduces to:

$$\frac{p_i - mc}{p_i} = -\frac{1}{\epsilon_{ii}}$$

This extreme case of no arbitrage shows that the degree of arbitrage only effects the matrix of elasticities, but does not alter the inverse elasticity and welfare results. If arbitrage is possible, then markets are no longer independent. Note that without arbitrage, the monopolist has a higher markup. It is therefore in his interest to prevent arbitrage.

3 Quantity discounts and quality premia

In general, a monopolist charges different prices in order to appropriate as much consumer surplus as possible. In order to be able to do this, he must know the different consumer willingnesses to pay. But what if willingness to pay is consumers' private information? Clearly, the monopolist especially would like the high valuation consumers to reveal their types, whereas precisely the high valuation consumers are the least willing to give up their private information. What should an optimizing monopolist do in this situation?

3.1 Quantity discounts

Assume there exist two types of consumers; L, who values quantity q at $V_L(q)$, and H, who values quantity q at $V_H(q)$. Furthermore, $V_L(0) = V_H(0) = 0$, which means that both value nothing at zero, and $V'_H(q) \geq V'_L(q) \ \forall \ q > 0$, which means that the high type is always wiling to pay more for an increase in quantity than the low type.

In the case of a quantity discount, different quantities are offered at different per unit prices. The monopolist offers two bundles (q_L, p_L) and (q_H, p_H), with p_i being the price for the quantity q_i, such that both types of consumers are willing to purchase, i.e., $V_i(q_i) - p_i \geq 0$, $i = L, H$, and such that both types choose the bundle that is targeted for their type, i.e., the bundles must be incentive compatible. This means that each type maximizes its utility by choosing the 'right' bundle, or

$$V_i(q_i) - p_i \geq V_i(q_j) - p_j,\ i = L, H,\ j \neq i$$

The monopolist is assumed to have constant marginal cost c, and to maximize profit $\Pi = p_L + p_H - c(q_L + q_H)$

The fact that the high type is always willing to pay more for a given bundle than the low type implies that the monopolist will always choose $q_L \leq q_H$. Otherwise, the high type would prefer the low type bundle, given the latter's lower valuation. The low type bundle will be such that the low type is just willing to purchase, i.e., $V_L(q_L) = p_L$. If his surplus would be strictly positive, the monopolist could raise the prices charged to both types with the same amount, thereby increasing his profit, without violating any constraint. If the low type is willing to buy, i.e., $V_L(q_L) \geq p_L$, and if the high type bundle is incentive compatible, i.e., $V_H(q_H) - p_H \geq V_H(q_L) - p_L$, then it follows that the high type is also willing to buy, for he values quantity higher. Furthermore, the high type bundle will be such that the high type's incentive constraint holds with equality. Otherwise, the monopolist could increase profits by charging a higher p_H.

This leads to the following conclusions:

1. The high type is offered the efficient quantity (i.e., $V'_H(q_H) = c$), whereas the low type gets strictly less than the efficient quantity.

2. The high type has a positive consumer surplus (if $q_L > 0$), whereas the low type gets zero consumer surplus.

The intuition behind this result is as follows. In trying to extract the high surplus of the high type, the monopolist faces the risk that the high type will choose the low type bundle. In order to prevent this from happening, the monopolist offers a relatively low quantity to the low type. Because the high type values quantity more than the low type, this will reduce the risk that the high type chooses the low type bundle. From the monopolist's point of view, the surplus extracted from the low type decreases, but this is more than offset by the increase in surplus extracted from the high type. [1]

Quantity discounts can thus be viewed as price discriminatory actions. By offering different quantities at different per unit prices, a monopolist tries to sell more to consumers who value quantity higher, and extract a larger part of consumer surplus.

3.2 Quality premia

The above analysis can also be interpreted as a model for quality premia. In this case, the monopolist offers a range of qualities to consumers who differ in their valuation of quality. Again, the monopolist offers the efficient quality to the high type and a less than efficient quality to the low type. The objective of this is to deter the high type from choosing the low quality.

Thus, by offering different qualities of the same good at different prices, a monopolist can segment the market. For instance, the different classes offered by railroads and airways can be explained by this model. But also more extreme applications can be seen in practice. It is often observed that a producer offers different qualities, even if the cost of offering the low quality is higher than the cost of offering the high quality. For instance, once a high quality product is produced, a portion of it is damaged or otherwise changed into an inferior product. Even if making the product inferior is costly, it may be optimal for a producer to pay this extra cost. As before, by doing this, he deters the high type to choose the low quality, and is able to appropriate extra surplus by serving the low valuation consumers.

3.3 Tie-ins

Tie-ins arise whenever a manufacturer requires the purchase of one product in order to purchase another product. A bank offering an indivisible array of services, a tour operator selling comprehensive vacation plans, and selling new cars with tires already included, are all examples of tie-ins. Tie-ins, or commodity bundling, can also be analyzed using the model introduced in this section. To see the similarity, consider a quantity discount. The

---

[1] The analysis can easily be extended to the case where there is a continuum of types with a known distribution.

price of two units of a good is less than twice the price of one unit. In a sense, the two units are bundled, since buying the two units separately involves additional costs. In the case of a tie-in, a producer offers a bundle of two different products together at a lower price than the sum of the individual prices. So, one can analyze the tying of units of several goods in a similar way as that of several units of the same good.

4 Peak-load pricing and Priority pricing

Several issues arise with respect to pricing if capacity is somehow limited. Two issues are peak-load pricing and priority pricing.

4.1 Peak-load pricing

The problem of peak-load pricing arises if a firm has two kinds of costs; capacity costs and a marginal production cost. The idea can be applied to many industries where building capacity is necessary to provide the product or service. Examples are pipelines, airlines, telephone networks, and electricity. In these industries, marginal cost pricing is not sustainable, since it does not take into account the price of capacity. Therefore, a capacity charge is necessary. Assuming that the capacity lasts for more than one period, peak-load pricing deals with the question to what period the charge should be allocated.

Consider a two period problem and let the subscript denote the different periods. The firm's profits are given by

$$\Pi = p_1 q_1 + p_2 q_2 - \beta \max[q_1, q_2] - mc(q_1 + q_2),$$

where β denotes the capacity charge. The problem is analyzed from the viewpoint of a social planner, [2] who will maximize

$$W = \int_0^{q_1} p_1(x)dx + \int_0^{q_2} p_2(x)dx - \beta \max[q_1, q_2] - mc(q_1 + q_2)$$

subject to a profit condition for the firm. Writing the Lagrangian and taking first order conditions gives

$$\frac{p_i(q_i) - \beta 1_{q_i \geq q_j} - mc}{p_i} = \frac{\lambda}{1+\lambda}\frac{1}{\epsilon_i}\ i = 1, 2,\ i \neq j$$

where $1_{q_i \geq q_j}$ is the characteristic function of the event $q_i \geq q_j$. If we let the demand in period 1 exceed the demand in period 2, two cases can be distinguished.

---

[2] It is assumed that demands are independent. This is not a necessary assumption; it simplifies the analysis, however.

- $q_1 > q_2$
  It follows directly from the first order conditions that all the capacity charge is allocated to period 1. With the capacity charge, the quantity in period 1 still exceeds the quantity in period 2. So, charging a higher price in the peak-load period does not change the peak. Moreover, note that it is socially optimal to charge a higher price in the peak-load period.

- $q_1 = q_2$
  Since the first derivative of $\beta \max[q_1, q_2]$ could be anything between 0 and β, the first order conditions become inequalities. Solving them for $q_1 = q_2 = q$ gives

  $$p_1(q) - mc + p_2(q) - mc = \beta$$

  This equation shows that the capacity charge is shared across the two periods proportional to the inverse demand, $p(q)$. Note that the charge is not shared proportional to elasticities, as was the case in previous applications.

4.2 Priority pricing

An issue that is also related to capacity, is the issue of priority pricing. Suppose that supply is stochastic, and that it is impossible to use prices to ration the market ex post, i.e. after the supply has realized. Priority pricing is a means of contracting in advance when capacity is stochastic.

There is a continuum of consumers, each desiring one unit of consumption. The consumers are ranked by their valuations for the good, so that the q^{th} consumer has a value $p(q)$ for the good, and p is downward sloping. Capacity is a random variable with distribution F. Now, priority pricing is a charge schedule c which provides a unit to a consumer paying $c(q)$ whenever realized capacity is q or greater. High valuation consumers, i.e. low q consumers, are willing to pay more to increase the probability that they will be provided a unit of the good.

The schedule has to be designed such that each type of consumer pays the charge targeted to him, i.e., it has to be incentive compatible. In other words, a consumer of type q should choose to pay $c(q)$ for the q^{th} spot in the priority list. So,

$$u(q) = (p(q) - c(q))(1 - F(q)) \geq (p(q) - c(\hat{q}))(1 - F(\hat{q})) \;\; \forall\, \hat{q} \neq q$$

Thus, for each type q, $c(q)$ must be chosen so as to maximize $u(q)$ with respect to q.

Note that the above story holds for the case of competitive supply. A monopolist might have an incentive to withhold capacity and charge higher prices. The monopolist can do so by offering another distribution of supply, G say, where $G \geq F$. It can be shown that, provided that marginal revenue is single-peaked, a monopolist will cut off the capacity at the monopoly supply, that is, at the point where marginal revenue equals zero.

5 Price dispersion

There is an important difference between price dispersion and price discrimination. Price dispersion results from imperfect information on the part of consumers, whereas price discrimination involves distinct prices for different types of consumer preferences. If all consumers would be perfectly informed, there would be no price dispersion, for all consumers would buy the good at the lowest price. Thus, different prices arise from different consumers having different information sets.

5.1 Search costs

If one assumes that acquiring information is costly to consumers, i.e., they have positive search costs, a reason for the differences in consumers' information sets may be differences in search costs. A consumer with low search costs will search longer, and may pay a lower price. However, a high search costs consumer may not find it worthwhile to continue searching, and pay the higher price.

The difficulty with this kind of models is what is known as the Diamond paradox. It can be illustrated as follows. Suppose that consumers all have a cost of search greater than some common lower bound γ, and that the search costs arise per store. It can be shown that in this model, all firms will charge the monopoly price, so the price dispersion disappears. The proof is by contradiction. Suppose that the lowest price is below the monopoly price, and consider a firm that sells at this price. It could raise its price by $\epsilon < \gamma$ without losing any customers, and thus increase profits. Therefore, a profit maximizing firm would never charge a price below the monopoly price.

5.2 Equilibrium price distributions

The old-fashioned way around this paradox was to assume differences between firms. However, this solution is not very attractive. The more modern models take another route. They assume that consumers have zero search costs or some other source of free information. By letting identical firms randomize, an equilibrium price distribution can be derived.

For example, consumers may get informative advertisements from sellers. Suppose n firms send advertisements, i.e., price offers, randomly assigned to a proportion α of the population. The consumers are passive; they buy at the lowest price offer they receive. First note that there is no equilibrium in pure strategies, that is, it cannot be an equilibrium for each firm to charge the same price. If this price gives all firms a positive profit, it would pay any single firm to undercut this price. If the price gives zero profit, any firm could earn strictly positive profits by offering a higher price, since there will be a fraction of the population that receives only his ad, and therefore buys at the higher price. So, in equilibrium, each firm will send out a random price offer.

Given that a consumer receives a firm's ad, and letting $F(p)$ be the probability that a price offer is not more than p, the firm's expected profit per consumer is

$$\pi(p) = (p - c)\left[1 - \alpha + \alpha(1 - F(p))\right]^{n-1},$$

since a consumer will buy from the firm instead of one of its competitors if he does not receive an ad from a competitor or if the competitor's offered price is higher. It is also assumed that all consumers have the same maximum price for which they are willing to buy the good, v say. Note that there must be at least some firm that offers v. Suppose not. Then the firm that sends out the highest offer could have sent out a higher price, without losing customers. We can calculate the expected profit of offering v as $\pi(v) = (v - c)[1 - \alpha]^{n-1}$, since $F(v) = 1$. The last expression must be equal to $\pi(p)$ since all firms must have equal expected profits in equilibrium. By setting $\pi(p) = \pi(v)$, a closed form solution for the equilibrium price distribution can be derived. For higher α, the distribution puts more weight on prices close to marginal cost, and the expected value of the best price realized by consumers is closer to the competitive level.

6 The Coase conjecture

Selling a durable good creates a problem for a monopolist. In the first period, it is optimal for him to sell at the monopoly price. After selling the monopoly quantity, the monopolist is tempted to sell a bit more at a lower price to lower valuation consumers, and a rational monopolist will indeed do this. But rational consumers, who face the choice in which period to buy, correctly anticipate the monopolist's behavior. Some consumers will thus prefer to postpone their purchase to a later period, i.e., the monopolist faces a different demand. Whether or not a consumer will hold out for future prices depends on his valuation of the product over time and the discount factor. However, when the time period between two price changes becomes arbitrarily small, all consumers will wait and buy at a lower price. Furthermore, the consumers know that the monopolist has an incentive to keep the periods between two price changes small, since this increases his profits. Therefore, if the monopolist can change prices sufficiently fast, the price must go to marginal cost arbitrarily quickly. So, the monopolist will price at marginal cost and loses his monopoly power completely.

The problem for the monopolist is that he cannot commit to follow a predetermined sequence of prices, although he would like to. In other words, it is as if the monopolist competes with himself. It can be shown that if commitment were possible, the monopolist could just earn the static monopoly profit. Because consumers realize that committing to a price sequence is not subgame perfect, monopoly profits are reduced to zero.

Although is may be serious problem for the monopolist, one should not conclude that it is impossible for him to make any profits. There are several ways in which the Coase problem can be evaded.

One of the solutions is that the consumers' beliefs sustain a supergame equilibrium. The consumers believe that the monopolist will stick to the equilibrium price path, and if they observe an out-of-equilibrium price, they believe the monopolist to play the Coase path. This will cause prices to drop to marginal costs immediately. Since this causes profits to fall dramatically, this threat is sufficient to sustain the equilibrium. Other ways to circumvent the problem of being his own competitor in future periods are to lease the goods instead of selling them, to give a money back guarantee, or to reduce the durability of the good, i.e., to turn the good into a nondurable good.

Computational Economics
Kenneth Judd

Report by Laurens Swinkels*

1 Introduction

For a long time, economists have been able to develop theories by mainly verbal reasoning. Gradually these theories were translated into mathematics and by logical reasoning the assumptions underlying these theories were uncovered. However, these models had to be kept simple in order to yield a closed-form solution. The derivation of such analytical solution is very difficult – or in many cases even impossible – in more realistic, and hence more complex, economic models. The easiest response to the non-existence of such solution is to stop doing research in such field. In recent years, however, a new branch of researchers has focused on finding numerical solutions to these complex problems.[1]

The increase in computer power has enabled the development of new optimization schemes which find numerical solutions for these complicated problems within a feasible time span. These newly developed solution methods are nowadays available in standard software packages. It usually takes some time to get familiar with the syntax used by these programs, but this is usually worth the effort for applied researchers. They should not waste their time by trying to write new computer code themselves, since these widely available optimization routines are programmed very efficiently.

The ability to find numerical solutions to difficult optimization problems has proven to be of much help in a wide area of economics, e.g. macro economics, public finance, and game theory. These lines of research typically used to resort to one or two period models in order to be able to find a solution, but nowadays dynamic multiple period models are

---

*Department of Econometrics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg.

[1] In recognition of the importance of this new field in economics, the Society for Computational Economics was founded in 1995 and now has an annual conference to exchange new research ideas. This conference is alternatingly held in Europe and the United States.

common and applied researchers are required to understand the basics of the computation behind their economic models.

This paper, which is based on the NAKE lecture series of Professor Kenneth Judd at the Free University in Amsterdam, tries to give some intuition and guidelines for researchers who have to solve analytically non-tractable optimization problems. In the next section, we will describe some of the basic numerical methods and their robustness. In Section 3, we focus on approximation methods for unknown functions. In Section 4, projection methods and its application in macro economics are discussed, and Section 5 describes perturbation methods. Finally, Section 6 concludes this report.

2 Numerical and Theoretical Solutions

For many simple problems, it is quite easy to find theoretical solutions. However, to find the numerical answer to the same question can be hard. Take for example an invertible matrix A and a vector b. If we want to solve the set of linear equations

$$Ax = b, \tag{1}$$

we know that this simply boils down to

$$x = A^{-1}b.$$

When we are actually faced with such problem of a sizeable dimension, we usually have to use the computer to solve it to get an numerical answer. Rounding errors in b may have a large impact on the estimated x. To investigate the robustness to rounding errors, we can develop an error model

$$A\left(x + e\right) = \left(b + r\right),$$

and define the elasticity of error by

$$\xi = \frac{\|e\|}{\|x\|} \div \frac{\|r\|}{\|b\|}, \tag{2}$$

with $\|\cdot\|$ the norm operator. The ξ is the ratio of the output to the input error. We can define the condition number of a matrix in order to approximate the elasticity of error ξ. The condition number is denoted by $cond(A)$ and defined as

$$cond(A) = \|A\| \left\| A^{-1} \right\|.$$

The condition number is a measure of determining whether matrix A is near-singular. It is more useful than the determinant of matrix A, which is mistakenly used by many researchers for this purpose. For example, a clearly non-singular matrix would be $B = \varepsilon I_n$, where I_n is the unit matrix of dimension n. The determinant of this matrix B would be near zero when ε is small (to be precise, $det(B) = \varepsilon^n$), but the condition number would be exactly equal to one. If the determinant of a matrix would be truly zero, then the condition number would be infinity.

In order to illustrate the fact that the influence of these round off errors can be large we constructed the following simulation example. First, a random square matrix A of dimension five is drawn, together with a five-dimensional random vector b. Then, we rounded all numbers of b to three decimals, and this vector is called b^*. We calculated both solution x and x^*, belonging to the vectors b and b^*, respectively. We then proceeded by calculating the norm of the differences in both solutions and the condition number of matrix A. This procedure is repeated thousand times. The results are plotted in Figure 1. On the horizontal axis we put the condition number, and on the vertical axis the norm of the rounding error. The results illustrate what was mentioned before, a rounding error may have a large influence on the results, especially when the condition number of a matrix is large.

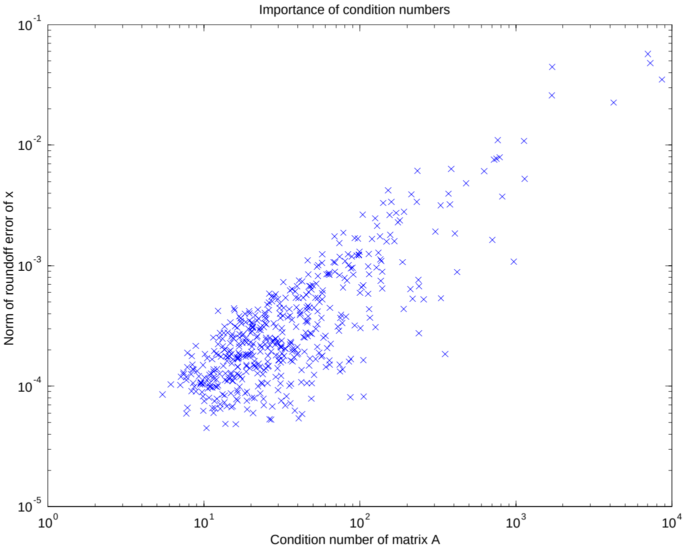

Figure 1: Error of calculation of $x = A^{-1}b$ in rounding off b to three decimals. On the horizontal axis is the condition number of the matrix, on the vertical axis the norm of the error in x after rounding off b.

These approximation errors might not just occur when we manually round numbers to three decimals, but also when machine precision is involved. Especially in sequences of optimization problems, these small errors add up and may become influential for the final result. For example, a sequence of linear problems arises when we use Newton's optimization method. Theoretically, the optimization problem $\min_x f(x)$ has a clear solution, i.e. $x^* = \arg\min_x f(x)$. However, to calculate the x^* that minimizes f can be problematic. A well-known method to solve this problem is Newton's method, which approximates the function f with a quadratic Taylor series expansion. If we denote J to be the gradient and H the Hessian of function f, the iteration scheme to find the optimal solution is

$$x^{k+1} = x^k - \left[H\left(x^k\right)\right]^{-1} J\left(x^k\right)'.$$

The iteration procedure starts by taking any x^0 as an initial guess, but the method converges faster when a sensible point near the optimum is chosen. The iteration procedure stops when a stopping criterion is met. This can either be the distance between two consecutive iterations, or the size of the gradient, which should be close to zero in all directions for the optimal solution. The main disadvantage of this method is that one may end up in local minima or maxima, and thus fails to identify the true global extreme value. If the Hessian matrix has a large condition number, this procedure generally converges slowly to the true value, if it converges at all.

3 Approximation Methods

Many applications require the estimation of a complicated unknown function. In traditional econometrics, we would for example think of a regression line through a large set of data points. The applications we discuss will not involve statistical regression techniques, but the links with regression will be mentioned at various points in this paper. The objective of this section is to construct an easy function $g(x)$ which is a good approximation of a difficult function $f(x)$, from which we have some data available.

Several issues may arise when trying to find this easy function. First, it is not immediately clear which points x are suitable to approximate f. Unlike with regression problems in which the data is given, we can choose which particular x we want to use for the approximation. A second issue is the shape of the function g, which might be depending on our knowledge about f. For example, it seems plausible that for functions with a periodic component other approximating functions are used than for functions without these cycle components. Third, an approximation without knowledge about its quality is useless, thus we want to know something about the distance between the function f and g. Connected with this is the question how easy an approximation g can be while still being relatively close to the original function f.

The challenge we face is to find algorithms that are reliable as more data becomes available. In this sense, we could speak of consistency of such algorithm, since the approximation should become closer when more information about the original function is available. These algorithms should be well-formulated in the sense that small computational errors should not heavily influence the resulting approximation. These ill-conditioned approximations appear to happen more often than one would like, so one should always be very cautious about this problem.

A straightforward way to approximate functions is to find a certain family of functions and fit some points from the original function f to this new function. The example we are going to elaborate upon here is the Lagrange polynomial interpolation. We have available data $(x_i, y_i)\,,\quad i = 1, 2, \ldots, n$. The objective is to find a polynomial of degree $n-1$ that agrees with all data points we have available. This polynomial of order $n-1$ is denoted by $p_n(x)$. When the points x_i from our data set are distinct, there is exactly one interpolating polynomial. In other words, there are no degrees of freedom left anymore. This is different from statistical regression problems, in which case we usually have many degrees of freedom and consequently the data points are close to, but not exactly on the regression line.

The advantage of using interpolation methods is that only a limited amount of data is necessary to calculate the approximating polynomial. The disadvantage of this technique is that the data has to be computed with very small error, because there are no degrees of freedom left after interpolation. One would expect that the approximating polynomial will get very close to the original function if the order of the polynomial is increased. In other words, we would like to have convergence between $p_n(x)$ and $f(x)$. An insightful example of the unreliability of this method can be found when investigating the function

$$f(x) = \frac{1}{1+x^2}$$

on the interval $[-5, 5]$. In the lecture notes the example is presented with 11-point interpolation; the figures presented here are for 5- and 13-point interpolation. It is clear from Figures 2 and 3 that the increase in the order of the polynomials does not increase the precision of the approximation.

These figures suggest that simple interpolation techniques are not very reliable and it seems in general not a very good idea to use these polynomials to approximate an unknown function. One may wonder what the reason is that this approximation is poor.

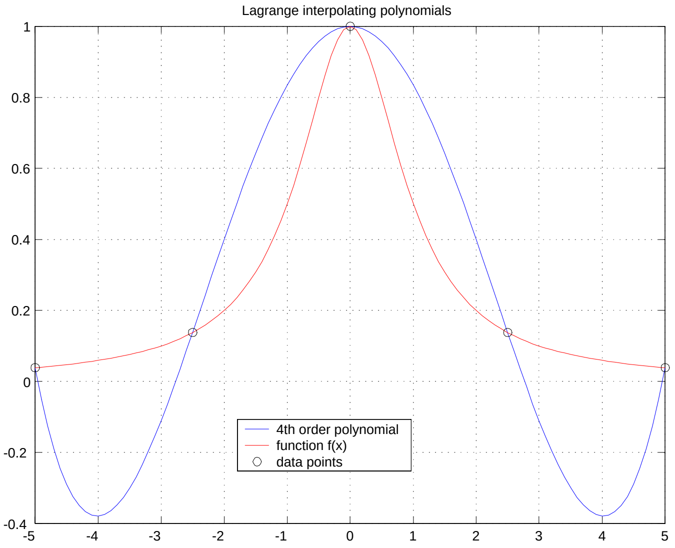

Figure 2: Approximating the function $f(x) = \frac{1}{1+x^2}$ with Lagrange interpolating polynomials. The set of interpolation points (denoted by little circles in the figure) equals one plus the degree of the polynomial and are equally spaced over the domain $[-5, 5]$.

Intuitively, it can be argued that the choice of the polynomials $1, x, x^2, ..., x^{n-1}$ is not particularly well-chosen. These vectors are highly correlated and therefore are not suitable to serve as a basis for the vector space of all polynomials. More appropriate bases would contain vectors that are orthogonal to each other. We observe the same phenomenon with classical regression. When the regressors are highly correlated, we observe that the point estimates are highly unreliable and we refer to this as the near-multicollinearity problem. The use of orthogonal regressors is to be preferred in classical regression to circumvent this problem, but there we can often not choose our observations ourselves. Since this is possible for the approximation methods, we prefer the use of orthogonal polynomials, since they generate reliable interpolation formulas. There are several types of polynomials that can be used, each one having its own merits. We decide to show the difference with the Lagrange interpolation of figures 2 and 3 and the Chebyshev interpolation, which are depicted in Figures 4 and 3. The approximation precision of a low order (Figures 2 and 4) are comparable. As we observed before, the use of higher-order polynomials with the Lagrange interpolation does not assure that the approximation gets closer to the true function. In the case of Chebyshev interpolation, we see that the approximation gets fairly accurate when the order of the polynomial is increased. The approximating function is

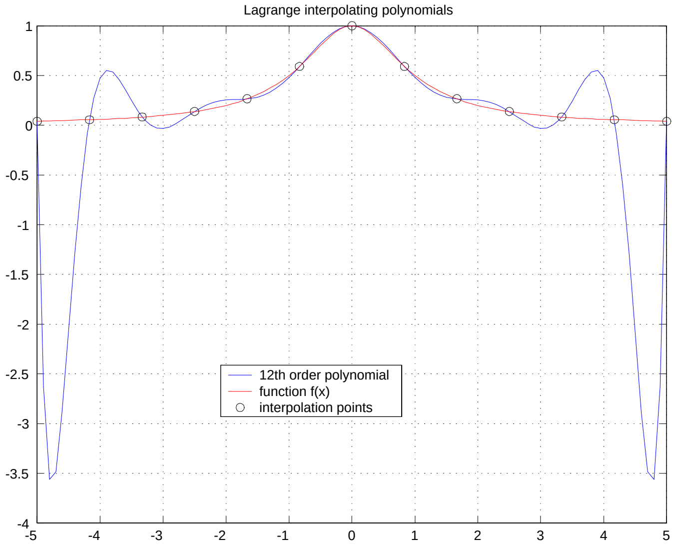

Figure 3: Approximating the function $f(x) = \frac{1}{1+x^2}$ with Lagrange interpolating polynomials. The set of interpolation points (denoted by little circles in the figure) equals one plus the degree of the polynomial and are equally spaced over the domain $[-5, 5]$.

rather wiggly around the true function, which indicates that the first derivative of both functions may be very far from each other. The dot near the point -4 has a positive slope for the true function, while the Chebyshev approximation has a negative slope. The quality of the slope estimate can be increased by doing a Chebyshev regression, which basically means that some degrees of freedom are needed by using some extra points in order to estimate the unknown coefficients.

Next to the Chebyshev polynomials, a variety of other functions can be used. Which one to use depends on the type of application one is working on. Other types are trigonometric, Legendre, or Hermite polynomials. Periodic functions may benefit from the use of trigonometric polynomials, while Hermite polynomials are related to problems which involve near normal distribution functions.

Other ways of approximation are for example rational functions, which are polynomials divided by other polynomials, or neural networks. One could also uses splines, which basically means that the domain is split up in several parts for which a polynomial is fitted separately. At the borders of the domains it is often required that both function values and first derivatives are equal, in order to keep a smooth function.

All the functions thus far suffer from the possibility that the shape of the data is not

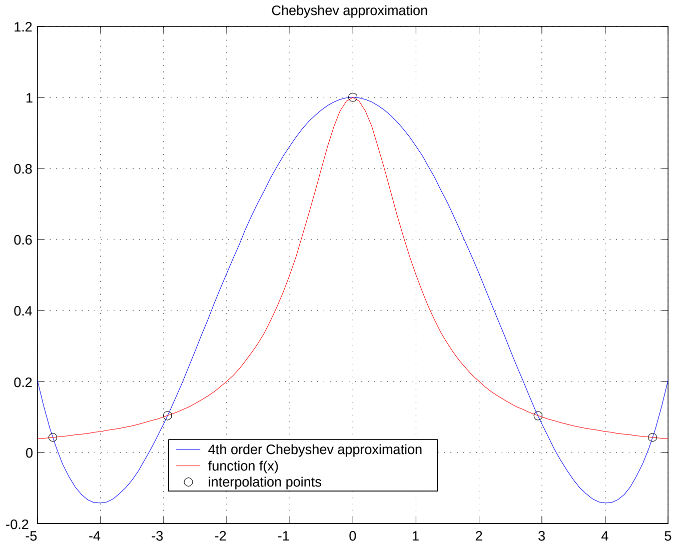

Figure 4: Approximating the function $f(x) = \frac{1}{1+x^2}$ with Chebyshev interpolating polynomials. The set of interpolation points (denoted by little circles in the figure) equals one plus the degree of the polynomial. The following recursion is used, $T_0(x) = 1$, $T_1(x) = x$, $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

preserved by the function estimate. When for example the points $f(x_i)$ are increasing for the x_i-s drawn, we might expect that the approximated function is also increasing. In general, this does not have to be the case, and one must be aware of this fact, especially for dynamic optimization problems. Some procedures can be followed in order to preserve shape, e.g. the Schumaker procedure for one-dimensional problems. For problems in a higher dimension things become more difficult, and even current state-of-the-art research on this topic has not yet revealed answers to all questions.

4 Projection Methods for Functional Problems

In the previous section, we analyzed how to approximate an unknown function. In the presented example, we actually knew the function which we wanted to analyze and observed how close the approximation was to the theoretical function. While such comparison is a very insightful exercise, for many realistic models the function we are after is really unknown and comparison with reality is impossible. Examples of the functions that economists want to estimate are consumption functions or pricing functions. The projection

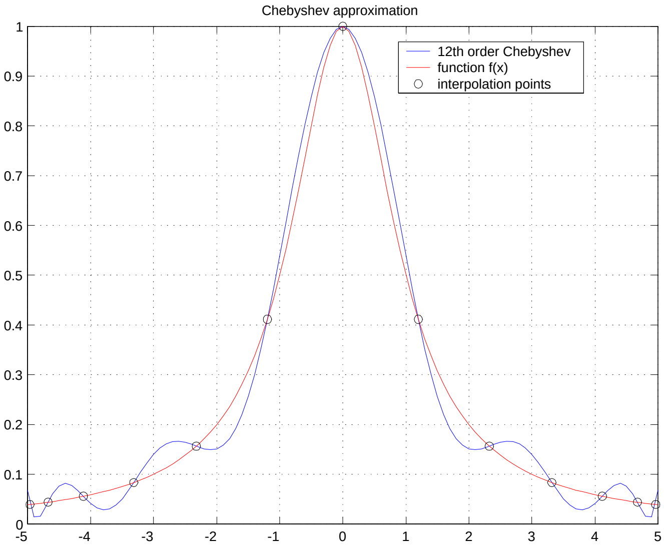

Figure 5: Approximating the function $f(x) = \frac{1}{1+x^2}$ with Chebyshev interpolating polynomials. The set of interpolation points (denoted by little circles in the figure) equals one plus the degree of the polynomial. The following recursion is used, $T_0(x) = 1$, $T_1(x) = x$, $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

method is a robust method for solving these type of dynamic optimization problems.

Consider the following example in which an agent wants to maximize her utility function by choosing her stream of consumption, and hence determines also the level of capital in the economy. The optimization problem is

$$\max_{c_t} \sum_{t=1}^{\infty} \rho^t U(c_t) \tag{3a}$$

$$k_{t+1} = f(k_t) - c_t, \tag{3b}$$

with U the utility function, c_t and k_t the level of consumption and capital in period t, f the production function, and ρ the time preference rate. If we derive the first-order conditions of this problem, it turns out that for optimality it must hold that

$$U'(c_t) = \rho U'(c_{t+1}) f'(k_{t+1}), \tag{4}$$

which implies that the ratio of marginal utilities in two consecutive periods equals the time preference factor times the marginal production rate. The problem is that we have an infinitely large number of unknown elements c_t, which form the solution. In order to find the optimal consumption path one can proceed by the following recipe. First, express the solution in terms of an unknown function. The consumption function can be written as $c_t \equiv C(k_t) = f(k_t) - k_{t+1}$ by observing the problem described in (3). The function $C(k)$ should satisfy equation (4)

$$0 = U'(C(k)) - \rho U'(C(f(k) - C(k)))\, f'(f(k) - C(k)),$$

and now define the non-trivial function on the right-hand side as $[N(C)](k)$, which is a function of k. In equilibrium, $N(C)$ should be equal to zero. The next step is to come up with an approximation for the function C, which we call $\widehat{C}$. This approximation might be a polynomial in k,

$$\widehat{C} = \sum_{i=1} a_i k^i, \tag{5}$$

which is according to some criterion close to the solution, i.e. $N(\widehat{C}) \approx 0$. We have now reduced an infinite dimensional problem to an n-dimensional problem without discretizing the state space form of the problem. Next, calculate the error that was made by the initial guess in equation (5), which we denote by R, and is a function of both the capital k as well as the coefficients of the approximation a,

$$R(k;a) = U'\left(\widehat{C}(k)\right) - \rho U'\left(\widehat{C}\left(f(k) - \widehat{C}(k)\right)\right) f'\left(f(k) - \widehat{C}(k)\right).$$

The final step of solving for $C(k)$ is to minimize the Euler equation errors R relative to a certain measure. The optimal solution can be obtained for example by minimization with respect to a of the squared residuals over all possible values of k. Another example is to use Galerkin, which uses n weighting functions to determine a. The method called collocation requires that the Euler errors are exactly zero for some prespecified values of k.

The measurement of the errors over the values of k requires an integration procedure. In general, exact integration is not possible and one has to resort to numerical integration. It is generally wise to use quadrature formulas (e.g. Chebyshev) to evaluate the integral numerically at sensible points. An alternative that is frequently used in practice is Monte Carlo simulation, but for high-dimensional problems this takes a tremendous amount of time.[2] Several large dimensional problems require the use of number theoretic methods to be solved. In order to solve for the unknown parameter vector a one can use Newton's method as long as the Jacobian is well-conditioned. If this is not the case one can use

---

[2] The disadvantages of using Monte Carlo techniques to solve these type of problems were emphasized repeatedly by Professor Judd, suggesting that many applied economists waste much time by refusing to use more advanced optimizations techniques discussed in this lecture series.

functional or time iteration methods, which are more complicated. The use of homotopy methods is advised only when nothing else seems to work, since its advantage of being almost surely globally convergent is offset by its complicated calculation.

In equation (5), the basis for the solution is taken to be the set $\{1, k, k^2, \dots\}$ which is in general not a very sensible choice for the basis of the polynomial space. The choice for Chebyshev polynomials might be more attractive in many applications that do not involve periodicities, for which Fourier functions are more suitable. In equation (5) the specification is chosen to be linear, which suffices for many applications. The use of neural networks, wavelets, and rational functions leads to incorporation of nonlinearities. Very little is known about the performance of these methods. Future research on this topic is necessary to provide sound recipes for problem solving.

5 Perturbation Methods

In this section, we will shortly describe the main idea of regular perturbation methods. The method boils down to approximating an unknown function by a Taylor series expansion around a trivial point of which the answer is known. By evaluation of the approximation over the area close to this known point the solution for the interesting cases are obtained.

Suppose we have a real-valued function h, which can be approximated by a Taylor series expansion at x_0

$$h(x) = h(x_0) + h'(x_0)(x - x_0) + \frac{1}{2}h''(x_0)(x - x_0)^2 + \dots$$

which can be computed by repeated differentiation of the function h. This approximation can be extended to vector-valued functions, but for expositional purposes this is not done here. The zeroth, second, and fourth order approximations of the function $h(x) = \frac{1}{1+x^2}$ at zero are

$$\begin{aligned} h^0(x) &= 1, \\ h^2(x) &= 1 - x^2, \\ h^4(x) &= 1 - x^2 + x^4, \end{aligned}$$

which are graphically shown in Figure 6. Note that the odd order approximations are equal to the even order approximations. This picture shows that the approximation close to the point in which the Taylor series is evaluated gets better when the order gets higher. This does not mean that the approximation it closer uniformly over the entire interval.

It can be seen in Figure 7 that close to zero the fourth order approximation is closest, but more than one unit from zero the zeroth order approximation is actually closest. We should be aware of the fact that approximations can only lead to sensible results if we stay close to the point around which the expansion took place.

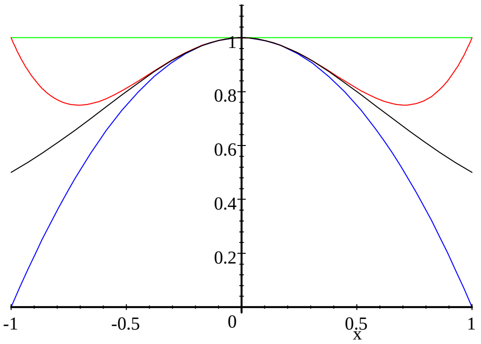

Figure 6: Taylor series approximations of the function $f(x) = \frac{1}{1+x^2}$. The function itself is shown in black, the zeroth, second, and fourth order appproximations are colored green, blue, and red, respectively.

The Taylor series are a numerical approximation technique, which is useful when h and its derivatives can be computed with ease at a point x_0. This is essential for the success of this technique. Note that for technical reasons the function h should be analytic in order to let this method work, but luckily this is frequently the case.

Suppose now that the function h is defined implicitly, which is the case in many applications, so that e.g.

$$H(x, h(x)) = 0. \tag{6}$$

If we know a special point (x_0, y_0) for which this implicit solution is true, i.e. $H(x_0, y_0) = 0$, the function h is known in a special point, i.e. $y_0 = h(x_0)$. If we derive implicitly equation (6) we obtain

$$H_1(x, h(x)) + H_2(x, h(x))\, h'(x) = 0$$

which implies

$$h'(x_0) = -H_2(x_0, y_0)^{-1} H_1(x_0, y_0)$$

for invertible $H_2(x, h(x))$.

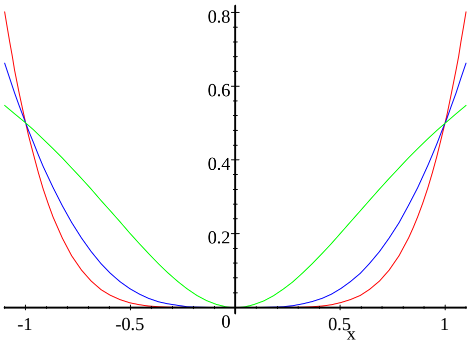

Figure 7: Errors of the Taylor series approximations of the function $f(x) = \frac{1}{1+x^2}$. The differences with the zeroth, second, and fourth order approximations are are colored green, blue, and red, respectively.

Now the function h can be linearly approximated by a Taylor series expansion around x_0, which yields

$$h^1(x) = h(x_0) + h'(x_0)(x - x_0).$$

In order to evaluate how close the linear approximation is to the solution we can calculate $H(x^*, h^1(x^*))$ and check its distance from zero, which it should be in equilibrium. We can impose a rule which states that the approximation is sufficient when this difference is smaller than a certain threshold value ε. A better approximation might be obtained when instead of the linear approximation the quadratic approximation is used.

The big advantage of using these methods is that researchers do not need to get bored by endless calculations with the possibility of human errors, but instead they can get reliable results very quickly by using widely available software packages.

6 Conclusions

During this workshop, we have become aware of the difficulties of actually calculating the numerical solution to a theoretical problem. In this report, we compare these numerical solutions to their theoretical counterparts to see how well they do. However, when the problems become more interesting, and hence more complicated the theoretical solutions cannot be computed anymore and we have to rely on the quality of the numerical approximation.

Numerical errors are important for calculation of the solutions, and sadly they cannot be avoided. All computers have some truncation point after which numbers are rounded, and these occur even in the simplest operations. When an algorithm to find a solution involves many computations, these individually small errors accumulate and may turn out to be a huge error in the final result. We should be aware of these problems and use methods that minimize the buildup of these errors as much as possible. For example, we could use a stopping criterion that allow iterative schemes to proceed until the round off error is relatively small. As illustrated, a matrix with a high condition number may lead to high errors in the solutions. Therefore, these ill-conditioned matrices should be avoided when solving linear problems. This is also observed when approximating a function with a polynomial. If we use a bad basis for the polynomial space, we are in fact using an ill-conditioned matrix and taking its inverse may result in numerical difficulties.

The methods used to solve problems that arise in economics are generally highly mathematical and economic intuition does not play a role to find the answers. We argue that this is good rather than bad. Most algorithms that arise by economic reasoning have very poor convergence properties, if they converge at all. We stress that in the problem formulation phase economic reasoning should be top priority, and motivations of the possibility to calculate the solution should stay at the background. However, when the formulation phase is finished, we should not try to find a solution that is motivated by economics, but resort to methods of numerical analysis who are especially designed to find a good approximation as fast as possible.

Nowadays, numerous commercial and non-commercial software packages are available to fully take advantage of the increased computer power to determine solutions to optimization problems. Since these sophisticated algorithms are programmed in fast languages, there is no need for the economist to program optimization routines. For them, it would be wiser to spend some time getting used to existing software (like e.g. CONOPT, NPSOL, GAUSS, or NAG) instead.

Innovation-based growth theory
Peter Howitt

Report by Egbert L.W. Jongen*

Introduction

Peter Howitt (Brown University), one of the leading researchers in the field of endogenous growth, gave a series of lectures at the NAKE-workshop at the Free University Amsterdam in June 2000. Below we report the main insights and ideas handed to us by Professor Howitt during these lectures. Peter Howitt and his continual collaborator Philippe Aghion view economic growth as a process of creative destruction, where leading innovators are continually displaced by the next leading innovator.

The outline of this report follows the outline of the lectures. In the first lecture we consider the limitations of 'exogenous' growth theory and the so-called AK-model of economic growth. In the second lecture we outline the bare bones model of innovation dirven by creative destruction. Lecture 3 takes a closer look at general purpose technologies, whereas Lecture 4 considers the topical issue of competition and growth. The final lecture, Lecture 5, seeks to explain the development of the world income distribution using the creative destruction approach of Aghion and Howitt, and deals with some criticisms of endogenous growth theory.

Lecture 1 - Exogenous growth and AK models

During the 1950s, Solow (1956) and Swan (1956) constructed the first general equilibrium models of economic growth.The striking implications of what came later to be known as 'exogenous' growth theory was that capital accumulation and population growth are

---

*Free University Amsterdam, Faculty of Economics and Econometrics, De Boelelaan 1105, NL-1081 HV Amsterdam, jongen@tinbergen.nl.

insufficient for a continual rise in per capita income, and policy can not affect the long-run per capita growth rate. Below we briefly consider the bare-bones of the Solow-Swan model of economic growth.

Let the production function be of the Cobb-Douglas type

$$Y(t) = K(t)^{\alpha}(A(t)L(t))^{1-\alpha}, \tag{1}$$

where $Y(t)$, $K(t)$, $L(t)$ and $A(t)$ denote aggregate output, the capital stock, labor supply and the level of technology, respectively. α is a production parameter. Note that labor in production equals labor supply, i.e. no unemployment, and technological progress is labor-augmenting. Labor supply and technology grow exogenously at the rates n and γ, respectively. Define $k(t) = K(t)/(A(t)L(t))$ and $y(t) = Y(t)/(A(t)L(t))$, where capital and output are now per effective unit of labor.The change in the stock of capital is endogenous, via the behavioral equation

$$\dot{k(t)}= sy(t) - (\delta + n + \gamma)k(t), \tag{2}$$

where the behavioral assumption is the constant saving rate s. δ denotes an exogenous depreciation rate. Note that n and γ effectively raise the depreciation rate of capital per efficiency unit of labor, as more effective labor units have to work on the same stock of capital accumulated in the past.

First, consider what happens when $n = \gamma = 0$. For low $k(t)$ the marginal productivity of capital (times the savings rate) is higher than the depreciation rate, whereas for a sufficiently high $k(t)$ the reverse is true. By the behavioral equation for capital accumulation this implies that there exists a unique steady state stock of capital k^* (next to the degenerate case $k^* = 0$). A higher savings rate or a lower depreciation rate will sustain a higher capital level, but output will cease to grow (or fall) once the capital stock settles down.

By introducing population growth, $n > 0$, output and capital can grow in line with the growth in labor supply, Following the same reasoning as before we find that the capital-labor ratio will eventually settle down at some k^*. At k^* $K(t)$ and $L(t)$ will grow at the same rate n, so output will grow at rate n as well. However, once again per capita growth can not be sustained.

When we introduce labor-augmenting technological change, $\gamma > 0$, the economy will again settle down at some k^*. Capital and output will now grow at the rate of effective labor supply $n+\gamma$. However, as labor supply in persons grows only at n, the economy can 'escape' from the presence of diminishing marginal product of capital. Per capita growth can be sustained. However, as γ is exogenous, policy can not affect the long-run growth rate of per capita income.

Another way to escape the ever present diminishing marginal product of capital is to eliminate this barrier to per capita growth directly, the so-called AK-approach to economic growth (Frankel (1962), Romer (1986)). In the AK-approach no labor augmenting technological change is required to sustain per capita growth. Suppose that aggregate output is given by

$$Y(t) = AK(t), \tag{3}$$

whereas the capital accumulation is the same as before. Effective labor supply is normalized to 1. One can readily show that capital and output will grow at the rate $sA - \delta$. Provided that $sA > \delta$ per capita income will grow without bound. Furthermore, by raising the savings rate policy can raise per capita income growth indefinitely.

Lecture 2 - Endogenous growth with creative destruction

The exogenous growth model could account for the 'stylized facts' of economic growth of Kaldor (1961) like i) sustained growth; ii) a constant capital-output ratio; iii) a constant labor share; and iv) the constancy of the interest rate on capital. However, the exogenous growth model does rather poorly in terms of more recent 'stylized facts' of economic growth. Furthermore, exogenous growth models leave the single engine of per capita growth, technological change, unexplained. Finally, no one is compensated for technological progress, whereas market forces seem to lie at the heart of the innovative process.

It would take quite some time before 'endogenous' growth entered the stage though. From 1970 to the early 1980s normal science continued, but no really new and interesting ideas on economic growth appeared (see Solow (2000)). However, the thesis of Paul Romer (published in 1986) sparked the research into economic growth models. During the late 1980s/early 1990s various seminal papers appeared in the leading journals that showed how technological change can be endogenized (i.e. Romer (1986), Lucas (1988), Grossman and Helpman (1991) and Aghion and Howitt (1992)). In endogenous growth models technological innovations are driven by the profit motive. Furthermore, policy re-enters the stage, as policies can affect the long-run growth rate. Indeed, policy plays a central role in endogenous growth theory, as with the endogenization of the growth rate came along various market failures. Below we consider the bare-bones version of the Aghion and Howitt (1992) model of economic growth.

Output is given by

$$Y(t) = A(t)(x(t))^{\alpha}, \tag{4}$$

where $A(t)$ denotes the technology level at time t. Let L be the (constant) labor force, and let n denote share of the labor force devoted to R&D. $L - n$ of labor units are devoted to producing the intermediate input x (one unit of labor generates one unit of x each period). One unit of labor devoted to R&D generates a innovation in $A(t)$ at rate λ. The stepsize of innovations is denoted by γ (> 1). Denote by $A(\tau)$ the level of technology after τ innovations. Free entry into R&D implies

$$w(\tau) = \lambda \frac{\pi(\tau + 1)}{r + \lambda n}, \tag{5}$$

where $w(\tau)$ denotes the wage rate after τ innovations, $\pi(\tau + 1)$ denotes the profits arising from the next innovation, and r denotes the interest rate. Profits follow from the monopoly price for units of x by the leading innovator ($w(\tau)/\alpha$). The wage rate has to equal the rate at which an innovation occurs times the discounted value of the profits associated with this innovation. Profits are discounted by the interest rate and the rate at which the incumbent is displaced by the next innovation, creative destruction occurs at rate λn (in equilibrium n will be independent of τ).

The equilibrium share of labor devoted to R&D is independent of the level of $A(t)$. Define the growth corrected wage rate by $\omega(\tau) = w(\tau)/A(\tau)$. n can be found from the free entry condition of R&D (with the right and left hand side divided by $A(\tau)$) given above and labor the market equilibrium condition

$$n + x(\omega) = L. \tag{6}$$

A higher share of labor devoted to R&D is associated with a higher 'wage rate' ω via the labor market clearing condition. From the free entry condition for R&D a higher share of labor devoted to R&D is associated with a lower ω (profits are discounted more heavily due to increased creative destruction). Equilibrium is a pair (n, ω) that satisfies both conditions.

One can readily verify that x is constant in equilibrium (as ω is constant). Hence, by standard growth accounting we have $g_Y = g_A$. From the poisson distribution with parameter λn we find $g_A = \lambda n \ln \gamma$. Note that a higher labor force will raise n and hence the growth rate (the model is 'scale dependent'), and so will a R&D subsidy (effectively lowering labor costs for R&D). Policy can influence the long-run growth rate. Howitt further showed how the model can be extended to allow for an infinitum of sectors and capital accumulation (see Chapter 3 in Aghion and Howitt (1998)). As a result of the continuum of sectors growth will loose its stochastic character at the aggregate level. Furthermore, capital accumulation becomes an essential ingredient for growth, as capital is one of the resources required for innovation. The endogenous growth results are unaffected, with a higher savings rate now also raising the growth rate of the economy.

Lecture 3 - General purpose technologies

The Schumpeterian view of endogenous technological change can be used to examine the response of the economy to the invention of general purpose technologies (GPTs). The Schumpeterian model creates an interesting link between growth and cycles. GPTs, like the steam engine and computers, seem to generate higher output in the long-run at the cost of a lower output in the short-run.

The first shot on the time path of an economy to the invention of a GPT comes from an adapted version of the Helpman and Trajtenberg (1994) model, where the adaptation is in the endogeneity of technological change. The arrival rate of a GPT is exogenous. Before the GPT can be used to produce (higher) output, components have to be discovered so as to facilitate its implementation. The discovery of components depends on the resources devoted to R&D. Output goes through three stages. In stage 1 all firms produce at the old GPT, with the components already discovered. Stage 2 begins with the arrival of a new GPT. Resources are directed towards R&D to generate the necessary components. Measured output falls. Stage 3 begins with the arrival of the necessary components. All firms produce with the higher GPT. Measured output rise above the level of stage 1. As the development of the necessary components is driven by the profit motive, the arrival rate of new GPTs has an ambiguous effect on the growth rate. A higher arrival rate implies more chances to upgrade to a better technology, but lowers the incentive to invest in R&D due to higher creative destruction.

The simple setup outlined above seems to have two main limitations: i) output jumps down when a GPT arrives, then jumps up when the necessary components are discovered, and remains stable in between jumps, this hardly seems in line with the data; and ii) the fall in output is due to the divertion of resources towards R&D, with a share of R&D of only 2.5 percent of output this seems insufficient to generate the more dramatic swings in observed output. Therefore we generalize (see Chapter 8 in Aghion and Howitt (1998)).

The first limitation can be met by extending the model towards a continuum of sectors where innovation is characterized by 'social learning'. Let n_0 be the share of sectors that produce at the old GPT, n_1 be the share of sectors that are performing R&D to adopt the new GPT, and n_2 the share of sectors that succeeded in adopting the new GPT. The crucial assumption is that sectors can learn from each other, in the sense that the rate at which a given sector starts doing R&D depends on the share of firms that succeeded in adopting the GPT. A typical time pattern for the share of n_0, n_1 and n_2 is given in Figure 1. At some exogenous rate sectors enter into R&D, a move from n_0 to n_1. Initially the share is low and rising slowly, but as sectors engage in R&D and start succeeding into adopting the new GPT, a move from n_1 to n_2, the sectors that are drawn into R&D starts to rise more steeply. The share of sectors performing R&D eventually falls again, as more and more sectors succeed in adopting the new GPT, the terminal state. The corresponding time path for output is given in Figure 2. With few sectors performing ('unproductive') R&D output initially falls, with the rise in R&D output falls more steeply, but output eventually rises above its initial level as more and more sectors adopt the new GPT.

The second limitation can be met by enhancing the adoption costs to the new GPT. As outlined in Chapter 8 of Aghion and Howitt (1998) the costs of adoption may go beyond the costs of resources devoted (directly) to R&D. One may envisage reallocation costs of labor, for example costly search by workers moving from firms with the old GPT to sectors with the new GPT. Furthermore, the new GPT may require a different capital stock than the capital stock currently in place, tailor-made for the old GPT. The obsolecence of capital increases the costs of adoption. A calibration with realistic parameter values for osbsolence may generate the observed dramatic initial drop in output.

Aghion and Howitt (1998, Chapter 8) and Aghion (2001) further consider the impact of a new GPT on the wage distribution. They argue that Schumpeterian growth theory can explain the productivity slowdown over recent decades and the wage compression during the 70s and the subsequent rise in the between- and within-dispersion in wages across skill-levels. The adoption to a new GPT (information technology) outperforms the explanations based on trade liberalization, exogenous skill-biased technological change and deunionization.

Lecture 4 - Competition and economic growth

The basic Schumpeterian model outlined in Lecture 2 predicts that more competition lowers growth. Indeed, one can readily verify that a higher elasticity of demand, an increase in α, lowers the profits from an innovation and hence R&D expenditures. However, this

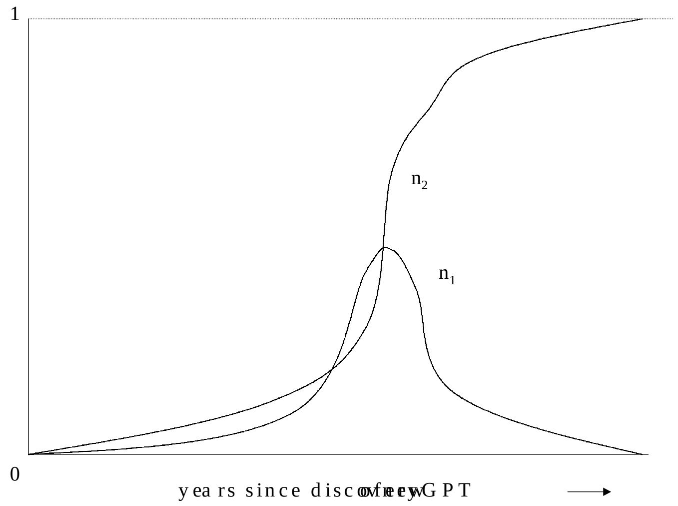

Figure 1: Shares of firms

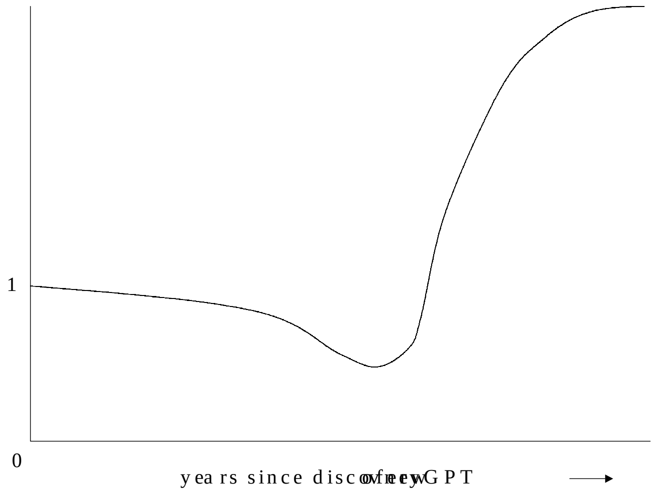

Figure 2: Measured output

seems to be at odds with the data (see e.g. Nickell (1996) and Blundell et al. (1995)). There are several ways to enrich the basic Schumpterian model to create a positive link between competition and economic growth.

First, more competition in the research sector may lead to higher growth. a) An entry cost for all firms to do research will lower the entry into R&D, a lower entry cost will stimulate R&D and hence growth. b) Suppose that R&D is done by a monopolist (asymmetric entry costs). The monopolist has a lower incentive to do R&D because his surplus (profits) rise only by the increment of technology, whereas an outside competitor would capture the increment of the surplus and the old surplus (the Arrow-effect). However, on the other hand the monopolist has a higher incentive to do R&D because an innovation raises the profitability of future R&D (a positive intertemporal spillover). Free entry into R&D may thus raise R&D and hence growth, depending on whether the Arrow-effect dominates the intertemporal spillover effect.[1]

Second, more competition in the product market may lead to higher growth. a) Increased competition may speed up the innovation process when the managers do not maximize profits but control rights (Aghion et al. (1997)). Higher competition increases the risk of being displaced. One way to avoid being displaced is to stay ahead, i.e. speed up the innovation process. b) An increase in the substitutability between old and new product lines will induce workers to move more speedily from the old to the new product line (the Lucas effect, see Aghion and Howitt (1996)). This induces a higher level of research and hence growth. c) Finally, neck-to-neck competition may foster growth (see Aghion and Howitt (1998, Chapter 7) and Aghion et al. (2001)). Following Howitt we consider this case in more detail.

The basic neck-to-neck model considers a sector with two competitors. The main difference with the basic Schumpeterian model is that sectors can be 'leveled' where both competitors have the same technology level, whereas they leapfrog in the basic Schumpeterian model. When a firm makes an innovation the competitor can costlessly acquire the previous technology. Hence, there are two cases to consider. The sector may

---

[1] It is sometimes argued that the Arrow-effect also exists in competitive markets (e.g. Barro and Sala-i-Martin (1995)). According to the author this is an error. The argument in favor is that an outside competitor captures the surplus of the incumbent and the additional surplus from the innovation, whereas the incumbent captures only the additional surplus from an innovation. However, free entry into R&D implies that total resources into R&D are fixed in equilibrium. Hence, an incumbent with rational expectations will realize that R&D will raise his or her probability of making an innovation and capture the additional surplus, and will reduce the probability that the incumbent loses his current surplus to an outside competitor. Hence, the incentive for R&D is the same for the incumbent and the outside competitor.

be unleveled, with one firm being one step beyond its competitor, or the sector may be leveled, with both firms having the same technology level. Due to the setup of the model both firms have an incentive to innovate when the sector is leveled, whereas the incumbent does not innovate when the sector is unleveled (the level of technology does not affect the level of innovation). Innovation costs are of a quadratic form. Hence, the same innovation rate can be achieved with less resources when both firms engage in R&D relative to the case where only one (the lagging) firm performs R&D. Leveled sectors perform more R&D than unleveled sectors. Higher product market competition lowers R&D expenditures in unleveled sectors (as in the basic Schumpeterian model). However, higher product market competition enhances R&D in leveled sectors. Hence, higher product market competition may stimulate growth. Note that there is an additional caveat as the rates at which sectors become leveled and unleveled are endogenous. Indeed, the model suggests that more product market competition enhances growth when competition is not too intense already. Aghion et al. (2001) further show (in a richer model) that some imitation will raise growth, whereas a lot of imitation reduces growth. Furthermore, the model predicts an important role for the complementarity between patent and anti-trust policies, with the impact of changes in the imitation rate and in product market competition being intimately linked.

Lecture 5 - Accounting for growth

The final lecture of Howitt dealt with the empirical challenges for Schumpeterian growth models. Mankiw et al. (1992) show that the standard Solow-Swan model seems to do rather well in explaining the cross-country distribution of per capita income growth, and the convergence rates of countries to their steady state growth rate. Furthermore, Jones (1995) shows that per capita growth shows no tendency to rise over a long time horizon, despite the rise in R&D expenditures. Finally, Evans (1996) shows that the developed countries seem to converge to the same growth rates, once again at odds with the basic Schumpeterian model which does not imply convergence. However, Howitt (2000) shows that an adapted Schumpeterian model of endogenous growth can meet these challenges.

First, consider the claim that endogenous growth theory seems to have little to add in explaining the distribution of per capita income growth rates and convergence rates over the Solow-Swan model. Recall from Lecture 1 above that according to the Solow-Swan model we have along the steady state path income per efficiency unit of labor as

$$y = \frac{Y}{AL} = k^{\alpha} = \left(\frac{\delta + n + \gamma}{s}\right)^{\frac{\alpha}{\alpha-1}}.$$

Taking logs, and adding a random error term ε_i where the index i denotes country i, we arrive at the following linear estimation equation of Mankiw et al. (1992)

$$\ln(Y/L)_i = \frac{\alpha}{\alpha - 1}(\ln s_i - \ln(\delta + g_{l,i} + \gamma)) + \ln A + \varepsilon_i, \tag{7}$$

where $g_{l,i}$ denotes the growth in the labor force (n_i denotes resources devoted to R&D by country i in the Schumpeterian model). The variations in labor force growth and savings rates lead to an $R^2 \simeq .75$. Not bad for a cross-section estimation. However, the share of capital in aggregate output (under perfect competition) α seems rather high at $\approx 2/3$. Linearizing the behavior of per capita income around the steady state yields a convergence rate of $(1-\alpha)(\delta + n_i + \gamma) \simeq .02$, which seems to accord with the data, for the same high value of α.

Howitt (2000) presents a hybrid Solow-Swan Aghion-Howitt model of per capita growth. As in the previous two lectures we will present the model in an informal manner. Countries produce output using a range of intermediate goods whose quality can be improved by engaging in R&D. A research success in country i in sector j occurs at rate $\lambda n_{j,i}$. A research success implies that the sector will jump to some exogenous frontier that grows at rate γ (at the end of the paper Howitt (2000) shows how the exogenous growth rate of the frontier can be endogenized by assuming $\gamma = \sum_{j=1}^{m} \sigma_j \lambda_j n_j$ where m denotes the number of countries). From this we can deduce that the growth rate of technology in country i is given by $\dot{A}_j\,(t) = \lambda n_{j,i}(A^{\max}(t) - A_j(t))$ where $A^{\max}(t)$ denotes the technological frontier. The level of R&D depends positively on capital per efficiency unit of labor (R&D is a fraction of output, hence capital enters into R&D) and on any (positive) subsidies to R&D. The capital intensity depends positively on the savings rate, hence R&D depends positively on the savings rate. Following the same steps as before we arrive at the following log-linear specification for per capita income for the adapted Schumpeterian model

$$\ln(Y/L)_i = \frac{\alpha}{\alpha - 1}(\ln s_i - \ln(\delta + g_{l,i} + \gamma)) + \frac{\kappa}{1-\kappa}\left(\ln\left(\frac{R\&D}{Y}\right)_i + g_{l,i}\right) + \text{constant} + \varepsilon_i, \tag{8}$$

where $\left(\frac{R\&D}{Y}\right)_i$ denotes the R&D intensity in country i. Hence, the Schumpeterian model yields an equation for per capita growth similar to the Solow-Swan model, but in addition suggests that R&D has explanatory power in per capita income growth. Lichtenberg (1993) shows that inclusion of the R&D variable reduces the estimate of α, the share of capital in income, which is desirable as noted above. Similarly, one may derive a convergence rate from the Schumpeterian model to the steady state level of output per efficiency unit of labor. A similar convergence rate to the Solow-Swan model can be achieved with a lower α, which once again seems to favor the Schumpeterian extention to economic growth. Furthermore, whereas the Solow-Swan model is silent on productivity differences across countries, these differences arise naturally in the Schumpeterian model. Indeed, countries that do less R&D will lag more behind the technological frontier than other countries.

The third criticism, Schumpeterian models do not lead to so-called 'club-convergence', i.e. the industrialized countries seem to have converged to the same growth rate, is invalidated in the model outlined above. Indeed, the law-of-motion governing the growth in a countries technology $\dot{A}_j\,(t) = \lambda n_{j,i}(A^{\mathsf{max}}(t) - A_j(t))$ suggests that the adapted Schumpeterian model does lead to 'club convergence'.

The adapted model can also meet the Jones (1995) criticism. In the adapted model labor force growth leads to an increase in varieties (sectors). However, with the appropriate choice of the aggregate production function (see Aghion and Howitt (1998), Chapter 12) one can counterbalance the implied rise of output by increasing varieties. Indeed, Aghion and Howitt (1998) argue that the gains from specialization do not obviously dominate any increase in complexity and errors that may occur as a result. This puts their model at the opposite extreme of Romer's (1990) expanding varieties model where all growth results from the increase in varieties. Consequently, R&D will rise, but its returns dissipate horizontally by the increase in product variety. The rise in R&D induced by population growth does not affect the growth rate.

Jones (1995) challenged the presence of endogenous growth on empirical grounds. However, Solow (2000) has challenged endogenous growth theory on theoretical grounds. Indeed, endogenous growth models typically need "a Santa Claus assumption to determine the growth rate endogenously." (Solow (2000), p. 105) An innovation in the basic Schumpeterian model raises $A(\tau)$ to $A(\tau + 1) = \gamma A(\tau)$, hence every improvement is assumed to lead to an equal proportionate increase in the current technology. This is the Santa Claus assumption. Indeed, when technology improvements are not in proportion to existing technology (for example $A(\tau + 1) = A(\tau) + \gamma$) per capita income growth will either fall to zero (as in the example) or explode as time goes to infinity. However, as noted by Howitt during his final lecture, even the Solow-Swan model needs to make a Santa Claus assumption, technological growth can only be purely labor-augmenting for the model to exhibit balanced growth. Hence, the arbitrariness in the technological process is a challenge for both endogenous and exogenous growth theory.

References

[1] Aghion, P., 2001, *Schumpeterian growth theory and the dynamics of income inequality*, mimeo.

[2] Aghion, P., Harris, C., Howitt, P. and J. Vickers, 2001, Competition, imitation and growth with step-by-step innovation, *Review of Economic Studies*, forthcoming.

[3] Aghion, P. and P. Howitt, 1992, A model of growth through creative destruction, *Econometrica*, 60, pp.323-351.

[4] Aghion, P. and P. Howitt, 1996, Research and development in the growth process, *Journal of Economic Growth*, 1, pp. 49-73.

[5] Aghion, P. and P. Howitt, 1997, A Schumpeterian perspective on competition and growth, in D. Kreps, ed., *Advances in Economics and Econometrics*, World Congress of Econometric Society.

[6] Aghion, P. and P. Howitt, 1998, *Endogenous growth theory*, Cambridge.

[7] Blundell, R., Griffith, R. and J. van Reenen, 1995, Dynamic count data models of technological innovation, *Economic Journal*, 105(429), pp. 333-344.

[8] Evans, P., 1996, Using cross-country variances to evaluate growth theories, *Journal of Economic Dynamics and Control*, 20, pp.1027-1049.

[9] Frankel, M., 1962, The production function in allocation and growth: A synthesis, *American Economic Review*, 52, pp. 995-1022.

[10] Grossman, G.M. and Helpman, E., 1991, *Innovation and growth in the global economy*, Cambridge.

[11] Helpman, E. and M. Trajtenberg, 1994, *A time to sow and a time to reap: Growth based on general purpose technologies*, Centre for Economic Research Policy, Working Paper no. 1080.

[12] Howitt, P., 2000, Endogenous growth and cross-country income differences, *American Economic Review*, 90, pp. 829-846.

[13] Jones, C.I., 1995, R&D based models of economic growth, Journal of Political Economy, 103, pp. 759-784.

[14] Kaldor, N., 1961, Capital accumulation and economic growth, in F.A. Lutz and D.C. Hague, eds., The theory of capital, London.

[15] Lichtenberg, 1993, R&D investment and international productivity differences, in H. Siebert, ed., Economic growth in the world economy, Tuebingen.

[16] Lucas, R.E., 1988, On the mechanics of economic development, Journal of Monetary Economics, 22(1), pp. 3-42.

[17] Mankiw, N.G., Romer, D. and D.N. Weil, 1992, A contribution to the empirics of economic growth, Quarterly Journal of Economics, 107(2), pp. 407-437.

[18] Nickell, S.J., 1996, Competition and corporate performance, Journal of Political Economy, 104(4), pp. 724-746.

[19] Romer, P.M., 1986, Increasing returns and long run growth, Journal of Political Economy, 94(5), pp. 1002-1037.

[20] Romer, P.M., 1990, Endogenous technological change, Journal of Political Economy, 98(5), pp. 71-102.

[21] Solow, R.M., 1956, A contribution to the theory of economic growth, Quarterly Journal of Economics, 70(1), pp. 65-94.

[22] Solow, R.M., 2000, Economic growth, 2nd ed., Oxford.

[23] Swan, T.M., 1956, Economic growth and capital accumulation, Economic Record, 32, pp. 334-361.