Trends in the Emergence of Agricultural Land Markets in Sub-Saharan Africa

Espen Sjaastad

1. Introduction
A critical difference between developed and less developed economies is found with regard to the extent and sophistication of market exchange (North, 1990). Hence, the recent emphasis on market integration and liberalisation in Africa, reflected in the structural adjustment and stabilisation programmes promoted by the World Bank and the International Monetary Fund. This emphasis extends to markets in land, evidenced by the promotion of private land rights and the associated increase in transactions that such privatisation is assumed to engender.¹

Objectives related to increasing transactions in land may, however, encounter obstacles linked to indigenous norms. Customary norms that prohibit land alienation – particularly sales, but also mortgages and leases – are pervasive in sub-Saharan Africa (hereafter SSA), and such norms do not necessarily break down when more individual rights of use and exclusion are introduced. Yet, quite recently, a number of publications have begun to catalogue burgeoning agricultural land markets across Africa.

Much has been written about why markets may evolve in SSA, and the supposed blessing or peril that they represent. Little, however, has been written about how such markets evolve. Bardhan (1989a: 7) notes more generally the persistent lack of accounts of

Note of Acknowledgements. I am grateful to Tor Arve Benjaminsen for his comments on an earlier draft and to the Norwegian Research Council for funding the research.

¹ Note, however, that privatisation of land, with its focus on the transfer of control over land from communities to individuals, is a somewhat different process from privatisation of state-owned (and often state-operated) enterprises.
how – rather than why – institutions change: ‘An institution’s mere function of serving the interests of potential beneficiaries is clearly inadequate in explaining it, just as it is an incompetent detective who tries to explain a murder mystery only by looking for the beneficiary, and on that basis alone proceeds to arrest the heir of a rich man who has been murdered.’

The purpose of this article is to explore various mechanisms and strategies that circumvent or undermine indigenous prohibitions against commercial land transactions in SSA. Such an exploration, though not exhaustive, may in turn point towards alternative or supplementary policies that will facilitate transactions when the blunt instruments of land reform are inadequate or inappropriate.

Section 2 provides a brief review of theory on why land markets evolve, how these are assumed to affect efficiency and equality, and gives an overview of indigenous constraints to market transactions in land. In Section 3, the mechanisms and strategies involved in overcoming these constraints are explored. Policy implications are discussed in Section 4.

2. Land Markets in SSA

Land markets, efficiency and equality
What is a land market? The distinction between commercial and non-commercial transactions in land mainly concerns whether or not parties engage in exchange, permanent or otherwise, with a view to mutual material gain. Thus, commercial transactions include a host of quite diverse exchanges, including sales, barter, mortgages and pledges, and various types of rental contracts. Traditional, non-commercial land allocation mechanisms in SSA also include both permanent land allocation – as governed by succession and gifts – and temporary land transfers in the form of borrowing, both types of transfer to some degree emphasising the recipient’s need for and ability to use farmland.

In the narrative of what Bromley (1989, 1991) calls the Property Rights School, shifts in relative prices drive institutional evolution. A shift towards higher explicit or implicit land values is generally believed to result from increasing population density, increasing agricultural commercialisation, or technological improvements. As land

---

2 Rental contracts include sharecropping, fixed rents, piece-rate wage contracts, pawn contracts and leasebacks.
values increase, institutional change becomes efficient, leading to ever more secure, specific and transferable land rights. Agricultural commercialisation may also create a demand for land transactions among individual households through increasing access to variable inputs, a general adjustment towards norms more tolerant of market transactions, and a shift to growing fewer crops or to monoculture, which in turn leads to a demand for more specialised types of land (see, for example, Demsetz, 1967; Johnson, 1972; North and Thomas, 1973; Ruttan and Hayami, 1984).

It is often assumed that formal titles are a prerequisite for the emergence of agricultural land markets (Platteau, 1996) but evidence, old and new, from a variety of African communities shows that this is not the case (see, for example, Brock, 1969; Platteau, 2000; Benjaminsen and Sjaastad, 2002; Mathieu et al., 2002). At the same time, title provision will not automatically lead to the emergence of land markets (see Platteau, 1996, 2000, and references therein).

The hypothesised benefits of land transfers are not fundamentally different from those of trade in any other asset or commodity: exchange permits transfers from less to more efficient users. In addition, liquidity will enhance asset value, since the right to convert land into other forms of wealth is valuable in itself. Land markets will also permit the use and acceptance of land as collateral for loans, a potentially important aspect in communities where alternative forms of collateral are scarce (Johnson, 1972; Platteau, 1996).

Land markets may also cause or accentuate wealth differentiation. Markets in land may cause greater inequality through the departure of inefficient farmers, concentrating land in the hands of the successful. In particular, farmers with high entrepreneurial and administrative skills will emerge as owners of land, whereas those without will emerge as labourers, a process that may be accelerated by structural characteristics of credit and insurance markets.³

Furthermore, an inverse relationship between farm size and productivity has been found in Malaysia (Bauer, 1946), India (Sen, 1962) and Japan (Okhawa, 1972). Insofar as this phenomenon extends to SSA, it would not only seem to offer a strong argument against land accumulation effects of markets; it would itself limit such

---

³ Because of administrative cost savings, it is cheaper for credit institutions to extend a few, large loans than many small loans, and credit institutions will tend to favour landowners with enough collateral to pursue large investments (Newbery, 1989). Moreover, Carter (1988) has established that, in the context of variable desirability of land as collateral and fixed supply of agricultural credit, privatisation of land rights may cause rationing effects, diverting capital from untitled farmers to titled farmers.
effects, since at some point farmers with large holdings would be unwilling to pay the reservation price of land demanded by smaller, more productive farmers. It is also an argument used by promoters of land reforms of the type where large, commercial farms are expropriated and distributed among small farmers (see, for example, Lipton, 1993).

Finally, it has been pointed out that whereas land markets may cause landlessness, the prospects of the landless availing themselves of a market in order to become farmers may be thwarted by the inclusion of collateral value and a real capital gains component in the price of land, causing prices to exceed the present value of net agricultural revenue streams (Binswanger et al., 1995).

The inequality effects of sales markets in land are less prevalent or are absent in rental markets (Fafchamps, 2000; de Janvry and Sadoulet, 2001; Lavigne Delville et al., 2001). The gap between exchange value and discounted net revenue flows will not exist in rental markets, since tenants will be unwilling to pay more than the reservation price; and the need for credit is usually absent, since tenants pay after, rather than before, use. Lower transaction costs may also give rental markets an efficiency advantage, though this may be offset by inferior investment incentives.

Given the above considerations, the emergence of land markets cannot be seen as unequivocally desirable but must be scrutinised, in each separate case, for the way the particular land transactions interact with other markets. On the one hand, under unfavourable conditions, land markets may lead directly to inefficiencies; on the other hand, they may contribute to increased inequality. And, as noted by Bardhan (1989a) and Platteau (1996), the concepts of equality and efficiency cannot be seen in isolation from one another.

Constraints on agricultural land transactions in SSA

To what extent is it justifiable to treat land tenure in SSA as a relatively coherent and ubiquitous set of rules? Simpson (1954: 51) held that, ‘In no aspect of native affairs can more diversity be found than in regard to the use and occupation of land…’ An initial question, however, is to what extent observed variations reflect fundamental differences in culture, or merely represent a differing maturity of progress along similar evolutionary paths. Simpson, despite the above emphasis on variety, seems to lean towards the latter view. White (1958), although acknowledging the existence of certain exceptions, shares this view. White also insists that differences in po-
Political structure do not vitally influence rules of access to land. It has also more recently been asserted that most tenures in SSA share a fairly broad range of basic characteristics (see, for example, Ault and Rutman, 1979; Migot-Adholla and Bruce, 1994; Platteau, 1996). Yudelman (1964), cited in Ault and Rutman (1979: 165), listed three cardinal principles of rights to land in Africa: there is no private ownership of land; security of tenure is guaranteed as long as tribal laws and customs are obeyed strictly; every member of the tribe is guaranteed the right to the use of land. Furthermore, a generally accepted proposition linked to indigenous land tenure in SSA is that rights to arable land are established through use; when land is cleared and crops are planted, rights to the land and the produce are removed from tribal or kinship control and become vested in the individual cultivator. When use is discontinued, the land reverts to the common pool. Significantly, permanent alienation of land is prohibited.

Some, however, would hold that many rules perceived as 'custom' in reality represent colonial legacies of manipulation and misinterpretation (see, for example, Amanor, 2002). Recent literature has unearthed a rich variety of rules and rights in SSA land tenures, while also stressing their flexible nature (see, for example, Ault and Rutman, 1979; Bates, 1984; Bruce, 1986, 1993; Migot-Adholla et al., 1991; Basset, 1993; Platteau, 1996; Sjaastad and Bromley, 1997). These writings emphasise the ability of indigenous institutions to respond to changes in external conditions such as population density and agricultural commercialisation.

Thus, while prohibitions against land sales may have been the norm, there is some evidence of increasing transactions in land in the region (Barrows and Roth, 1990; Lawry, 1993; Troutt, 1994; Platteau, 1996; André and Platteau, 1998; Lavigne Delville et al., 2001; André, 2002; Benjamin and Sjaastad, 2002; Mathieu et al., 2002). Migot-Adholla et al. (1991), for example, found that around 16 per cent of smallholder plots surveyed in six countries in the region had been acquired through the market mechanism.

In general, there would seem to be more activity in densely populated regions, supporting the tenets of the Property Rights School. But the evidence is by no means clear-cut; strong prohibitions against trade in land remain in many land-scarce areas across the continent. A transition towards increasing land transactions is therefore not an

---

4 Mainly seen as a question of whether political authority is vested in a tribal entity, with a chief at the apex and an associated conception of tribal territory, or in smaller groups defined in terms of kinship and marital affinity.
automatic consequence of changes in certain external conditions. A number of factors work to constrain agricultural land markets in general and in SSA in particular.

Factors that restrict development of markets in general will also constrain markets in land, such as dispersed populations, associated self-reliance and uniform natural environments that curtail specialisation. Distortions in other markets may themselves restrict transactions in land; the absence of a credit market will inhibit the financing of land purchases; the absence of a labour market may render the acquisition of more land meaningless. Moreover, the information costs attached to land are high, due to the complexity and variability of land and the effort required to understand this complexity. Partly for these reasons alone, land markets may be stuck in low-turnover equilibria.

Several factors specific to cultures in SSA may further impede land transactions. Markets in land under ‘communal tenures’ are unlikely to evolve since land, in the territorial sense, is not the object over which individual rights extend (Sjaastad and Bromley, 2000). Also, the significance attached to land in most communities goes far beyond the simple notion of a factor of production and its associated scarcity and price. Land takes on important functions of identity, security, and prestige related to kinship, culture and religion. These functions mean that land-holders are far more reluctant to part with land than they would have been if land were merely a productive asset, and this may in turn raise the relative price of land.5

Throughout much of SSA, tribal affiliation is an essential de facto determinant of the feasible area of residence for a farmer, thus in important ways restricting entry into whatever land market may exist. Resettlement – at least to areas desirable to the household itself – is often limited to what may take place within the boundaries drawn up by restrictive rules of kinship, affinity and clan membership. Moreover, as noted, custom often prohibits the sale of land, on the grounds that land is a common heritage of the kinship group or tribe, and land is often needed as collective security, for example, for urban labourers. Permanent transfers have traditionally been restricted

---

5 Basu mentions the frequent lack of interim transactions – the sale of an asset with an associated intent to repurchase a similar asset – as a reason for thin land markets in developing countries: ‘individuals hesitate to sell land because turnover is low; and it is their hesitation which, in turn, reinforces the low turnover’ (Basu, 1986: 164). Furthermore, ‘land sales are few precisely because land is the last asset people part with’ (ibid.: 172). Thus, non-economic aspects may lead to sluggish land markets, and sluggishness breeds more sluggishness.
to inheritance, bequests and *inter vivos* gifts.  
Finally, in the region considered here, institutional structures are often incapable of providing the security necessary for extensive and impersonal exchange of goods of substantial value (Bromley and Chavas, 1989). Markets for land will remain limited as long as enforcement mechanisms remain inadequate.

### 3. How Land Markets Emerge

An increase in land market activity may be facilitated simply through changes in technology, land use or culture. When permanent cultivation methods supplant shifting cultivation and forest resource extraction, subsidiary rights may disappear quite painlessly, and emerging territorial rights may include rights to alienate (Shipton, 1989; Sjaastad and Bromley, 2000). Where previously no rights to land existed, exogenous changes may lead to spontaneous appropriation and subsequent land market activity, as appears to have occurred among the Kipsigis of Kenya (Bohannan and Dalton, 1962; Manners, 1962). On the periphery of rapidly expanding urban centres, the brute and combined forces of local government and economic opportunity may quickly do away with resistance to sales, while the spread of religions that do not militate against trade in land – may also be an influence (Benjaminsen and Sjaastad, 2002). In general, however, agricultural land markets will emerge gradually by way of more subtle mechanisms.

**Reciprocity**

Perhaps the most obvious mechanism through which markets in agricultural land can emerge is through adaptations in traditional modes of transfer. Polanyi (1957) distinguished between three modes of transfer: market exchange, redistribution and reciprocity. Redistribution represents vertical transfers between traditional leaders and subjects – similar in principle to taxation and provision of public services – and is therefore unlikely to transmute into market exchange. But reciprocity (also described as gift exchange, solidarity

---

6 Shipton (1989) and Troutt (1994) note that previously active land markets may subside under very high population densities, since a minimum holding is necessary for maintenance of social identity and because efficient redistribution has been exhausted. Shipton also notes that colonial authorities, striving for control in rural areas, often worked to cement traditional prohibitions against land alienation.
networks or extended family mechanisms), which denotes lateral transfers without immediate or guaranteed compensation, offers possibilities. Reciprocal mechanisms are pervasive throughout SSA, though often weaker in areas where markets are advanced (see, for example, various writings in Bohannan and Dalton, 1962; Dalton, 1967; Gerard-Varet et al., 2000).

Reciprocity, as conventionally seen, consists of the extending of gifts from a household possessing an excess of a good to one experiencing a shortage, with the understanding that the favour will be returned if roles are reversed in the future. In the absence of exogenously determined prices, agents contemplate the credibility and performance of individual transactors, and seek to develop trust and assurance through repeated dealings. Reciprocal transfers tend to be communal rather than associational; whereas market exchange results in temporary associations serving some specific purpose, reciprocity serves to confirm and strengthen communal relationships already defined through social organisation (Dalton, 1962). Hyden (1990) also notes that whereas market exchange combines the presence of a *quid pro quo* and simultaneity, reciprocity is defined by their combined absence.

From an economic perspective, reciprocal transfers may have three distinct functions. First, reciprocity may act as a substitute for more conventional commodity markets, whose purpose is to exploit economies of labour division and specialisation (Kranton, 1996). Second, it may act to ensure an equal distribution of wealth (see, for example, Hyden, 1990). Third, a reciprocal transfer may represent a purchase of insurance (Posner, 1980; Bromley and Chavas, 1989; Fafchamps, 1992; Coate and Ravallion, 1993). The risk-spreading rationale has received most emphasis in economic treatments of reciprocity; in the absence of cheap insurance options, local communities pool their resources at the expense of the adverse incentives that such pooling involves. As noted by Fafchamps (1992), however, *ex ante* transfers of factors such as land can greatly reduce the necessity and adverse effects of *ex post* transfers of consumer items.

In the classic anthropological view of things, reciprocity is symptomatic of an entirely different set of values than what typifies societies dominated by market exchange. In the somewhat stereotypical view of Popkin (1979), anthropological literature in general – and Scott (1976) in particular – tended to glorify traditional arrangements of risk-sharing and to imbue pre-capitalist societies with a superior ethical code. Providing numerous examples of opportunistic behaviour in peasant communities, Popkin argued that traditional institu-
tions were designed to maintain rather than combat wealth and income differentials, and that risk aversion found expression in self-insurance rather than collective strategies. Over the last two decades, however, publications such as those of Posner (1980), Sugden (1986), Platteau (1991b), Fafchamps (1992), Coate and Ravallion (1993) and Gerard-Varet et al. (2000) have shown the conflict to be overstated. The idea that poor people in risky environments develop collective insurance mechanisms need not rely on altruism or different ethics.

Moreover, the distinction between reciprocity and market exchange, often regarded as clear-cut in the earlier literature, is no longer considered unproblematic. Commonly identified spaces in which the two forms of transaction were assumed to differ include the prior interdependence between the parties to the transaction, the alienability of the transacted good, the presence or absence of any immediate compensation, the degree to which such compensation is subject to precise calculation, and the ‘moral evaluation’ of the transaction. Increasingly, however, it is recognised that most of these spaces are continuous rather than binary, that a whole family of different transaction categories exists, and that each may possess any single property to a greater or lesser degree.7

These continuums provide opportunities for a gradual transition from pure reciprocity towards market exchange. For example, transfers may increasingly become subject to verbal guarantees of a more precisely calculated return transfer in the future. Or more immediate compensation may be demanded, and the range of goods regarded as acceptable return transfers may broaden; the ‘multicentric spheres’ described by Dalton (1962), whereby modes of transfer and the goods that circulate within them are sharply distinguished, may disintegrate. As noted by Bohannan and Dalton (1962), the introduction of money as a universal medium of exchange will tend to erase the boundaries between different spheres.

Such a transition is reflected in those communities where land transactions frequently contain elements of both trade and reciprocity. The distinction between loans and pledges may often be quite spurious; loans will often be attended by token return gifts, sometimes repeated annually, providing the lender with symbolic evidence

7 Recent literature has provided numerous examples of trade that is both personalised and supportive of social interrelations (see Geertz, 1978, 1979, for a well-known example), and Parry (1989) provides examples of gift exchanges that are subject to the moral condemnation often attributed to commerce.
of ownership. As land becomes scarce, such gifts may – gradually, imperceptibly – acquire a more tangible value (Lavigne Delville et al., 2001). Among the Arusha, for example, land pledges between relatives or friends involve an immediate cash transaction, but the average payment is considerably lower than what could have been obtained in an open market. As the demand for land increases, the cash component will rise and may eventually and seamlessly approach market price. Gibson (1962) noted that bridewealth among the Herero in South-West Africa also possessed these dual characteristics: part symbol, part trade. Gray (1962) in Kenya, and Shipton (1989) more generally, also note that when pledges are irredeemable upon the death of the pledgor, temporary transfers are converted into permanent ones.

Thus, a transition from reciprocity to commerce will often represent a relatively organic evolution of institutions. As resource scarcity changes, the benefits to be derived through challenges to the old order, and the cost of resisting them, will also be altered. Challenges may thus gain in vigour and number, and the incentive to penalise and thwart such challenges may simultaneously be reduced. Successful challenges create precedence for others to follow, causing a gradual and continuous process of validation and revalidation of new norms and principles; at some point, the old institutions may be abandoned entirely (Holy, 1986). Shipton (1989) notes how prohibitions against land sales have frequently been tied to curses and taboos; as traditional beliefs are abandoned and replaced by, for example, Christianity or Islam, so the incentive to abide by prohibitions will be reduced. Moreover, as observed in Tanzania (Young and Fosbrooke, 1960; Moore, 1986), Kenya (Sorrenson, 1967), Uganda (La Fontaine, 1979) and Rwanda (André and Platteau, 1998), once prohibitions are violated they may be irrecoverable; on land acquired by other means than inheritance, customary norms of redistribution and solidarity cease to apply.

More abrupt transitions may also take place. In Zambia, blatant examples of subverting traditional prohibitions against land alienation involved the simple insistence of the parties to the transaction that the land and money exchanged were in the form of gifts and counter gifts rather than trade (Sjaastad, 1998). Disguised sales of a similar type were observed by Carlsen (1980) in Kenya, and by Moore (1986) in Tanzania. As noted by Shipton (1989: 13) on sub-Saharan Africa in general, ‘… it is hard to control sales. Where land was scarce and in demand, farmers disguised sales as loans, rentals, or pledges’.
Another way in which the transition from reciprocity to commerce manifests itself is in the initiation of a reciprocal transfer (Kolm, 2000). Conventionally, a prospective recipient who signals a shortage in a specific commodity to potential donors initiates a reciprocal transfer. Increasingly, however, households and individuals in rural African communities will offer their services in anticipation of such shortages. Typically, households with excess labour resources will offer labour assistance during bottlenecks to households with labour shortages, with the expectation that land will be transferred in the other direction prior to the next crop season.

Permanent gifts and irredeemable quasi-pledges and loans notwithstanding, reciprocal transfers of land in rural Africa are traditionally of the temporary type. Moreover, for reasons to do with rights appropriation and security, land can most often be borrowed for only one or a few seasons at a time. This phenomenon – the fear that temporary transfers may turn into permanent ones – may also persist when reciprocal transfers begin more closely to resemble market transactions. Land markets that evolve through a gradual adaptation of reciprocal mechanisms may therefore initially be characterised by short-term rent arrangements, and this may in turn offer a possible explanation for the proliferation of tenancy arrangements in some African economies.

Finally, the networks within which reciprocal transfers take place are limited. Bromley and Chavas (1989: 720), in regard to semi-arid regions, state that ‘It is our fundamental premise that sustained agricultural development within the agro-ecosystem here being discussed is likely to occur only when economic agents have the opportunity to engage in a variety of economic transactions that transcend the traditional transactions among family and clan’. For trade in land to extend beyond these spheres, third-party enforcement of contracts is necessary.

Investments

Land rights appropriation in rural Africa has traditionally been linked to occupancy and cultivation (see, for example, Ault and Rutman, 1979; Cohen, 1980; Platteau, 1996; Sjaastad and Bromley, 1997). The key to the acquisition of rights over a natural resource is the labour expended on its transformation or extraction; through the mixing of labour with land, rights emerge. In areas where shifting cultivation dominated, for example, the felling or pruning of trees and subsequent crop cultivation established temporary land rights,
which would vanish when the crop rotation gave way to an extended fallow.

An investment represents a commitment to long-term use of and interest in land, and land improvement is thus potentially a powerful appropriation mechanism (Besley, 1995; Sjaastad and Bromley, 1997). It is, however, notable that individual or household rights could often be seen to extend over rights to the improvement itself rather than the land on which it is made (Cohen, 1980), a notion that has also been reflected in post-independence land legislation in SSA.

A tenure structure whereby rights are relinquished when use ceases prevents land accumulation and ensures a rough balance between the land and labour endowments of individual households as long as land remains abundant in the community as a whole. When rights are linked to use, however, increasing permanence of land use (caused, for example, by a combination of population growth and technological change) will engender increasing permanence of land rights. The distributive effect of fallow land re-entering the public domain is gradually lost, and a demand for alternative methods of redistribution appears. This demand will further increase if traditional patterns of succession and reciprocal transfer mechanisms disintegrate, and more permanent land use may thus in itself pave the way for the emergence of land markets. Demand evolves in response to changes in factor balances created by unequal access to resources and changes in household composition.

The above relates mainly to why, rather than how, land markets emerge. In addition, however, as land use becomes more permanent, multiple rights characterised by the ‘secondary’ claims of numerous households to a variety of natural resources such as fruits and firewood lose relevance (Sjaastad and Bromley, 2000). An increasing share of the complete bundle of land rights will thus be vested in a single household, and if the right to trade in land is dependent on such a share of rights – a high share is perhaps a necessary although not sufficient condition – then this process will facilitate the emergence of markets.

The norms that inhibit land sales do not commonly extend to investments affixed to the land (including unharvested crops), due to the essentially ‘human’ rather than ‘God-given’ nature of investments as opposed to the land itself (see Allan, 1965, or Cohen, 1980, for a more general statement of this proposition). Thus, when one invests in land improvements and these investments can be alienated, this will also promote a land market; although the value of the land itself cannot explicitly be transacted, there is no way to enforce such a rule
unless the improvements can be physically detached from the land or assessors become involved. If the transfers of land and improvements are interlocked, then the value of the land will inevitably enter into these transactions, subsumed by the indivisibility of the two goods themselves or the rights extending over these goods. As land becomes scarcer, the price will to an increasing degree reflect land value rather than the value of the improvement; farmers will increasingly purchase improvements in order to obtain land rather than the improvement itself.

Throughout much of rural Africa, for example, norms stipulate that whereas village land cannot be sold, dwellings can (Shipton, 1989). And the custom is that the land around the house, which may involve sizable permanent fields, is attached to the dwelling. A study of village conflicts and court cases in the Southern and Northern Provinces of Zambia reveals that this type of consideration also applies to permanent structures linked to irrigation and soil conservation (Sjaastad, 1998). With respect to generation of the right to alienate surrounding land through tree planting, Parkin (1972) found such linkages in Kenya, Middleton (1961) observed a similar process at work in Zanzibar, and Besley (1995) in Ghana. Investments therefore not only increase the duration and security of land rights, but may also facilitate land alienation through the inseparability of land and improvements. The important connotation here is that investments in land may be an important engine in any drive to establish markets in land; usually only the opposite process is acknowledged.

Moreover, Le Bris (1979), in Togo, observed that land could only be exchanged for other productive assets, not for consumer goods; the implication is that increasing market activity for assets in general may promote a land market.

Investments may, of course, affect alienation of land in another manner. Squatters, borrowers or tenants may find it profitable to undertake fixed improvements not so much to enhance production as to establish permanent claims. Farmers in SSA frequently express the opinion that ‘lending is the same as giving’ and that ‘renting is the same as buying’, and these fears are indeed the reason why it is often forbidden for borrowers or tenants to undertake investments (Mathieu et al., 2002). Given the complexity of kinship and land relations, however, as well as problems of proximity, it is not always a simple matter for an owner to enforce this rule. Concerns over adverse claims based on investments may, in fact, in part explain the absence of rental markets in certain areas (Benjaminsen and Sjaastad, 2002), and in such cases causality is reversed; rather than in-
vestments leading to increased market activity, fear of their connotations may stifle it.

Migration and caretakers
An apparent anomaly that may emerge from surveys of land transactions in rural Africa, and elsewhere, is that the number of households that have purchased land significantly outnumbers those that have sold land (Rao, 1972; Sjaastad, 1998). This might reflect reluctance on the part of responders to admit to what is perceived as a violation of traditional rules. It is, however, probably more a function of the rules of entry and exit that govern traditional access to land. Whereas the notion of outsiders buying land in the village is anathema to most traditional communities, there is still the problem of what to do with land left behind by villagers who emigrate to town or who decide to establish farms elsewhere, especially if these do not leave behind close relatives able to use the land. Thus, while entry is restricted, exit is not, and while buyers of land remain in the village, sellers have more commonly migrated, to urban centres, other farming communities or isolated and previously unpopulated areas.

The problem concerns the market dynamics that accompany a general migration of labour from rural to urban areas. In Africa, it is not so much a case of the depopulation of the countryside but a relative discrepancy in the population growth in rural and urban areas, caused by net urban migration and clustering. The particular mechanism in operation involves ‘caretakers’, who agree to use, manage and protect the land of émigrés for the duration of their absence from the village. The nature of such transactions may change fundamentally as land becomes increasingly scarce, the nature of migration itself changes, and communal or kinship control of rights to land also changes. Under conditions of relative land abundance and strong communal control, migrants will neither relinquish nor acquire rights (no transaction occurs), land simply reverting to the community upon their exit with the understanding that the right to cultivate some land is retained and can be exercised upon their return (Boserup, 1965; Ault and Rutman, 1979).

As land becomes scarce and rights presumably become more individual, however, the mechanism may increasingly acquire the nature of a transaction. As noted by Shipton (1989), people who migrate from the village in search of work, leaving their holdings in charge of kin, may also cede certain rights to their land in doing so;
rights, moreover, that migrants might be unable to reclaim upon their return to the village. Thus, Lawry (1993), in Lesotho, observed that ‘… data presented below suggest that increasing numbers of landless are middle-aged, returned mine workers’. Similarly, André and Platteau (1998), in the context of a Rwandan community characterised by increasing exclusion, note how it has become more difficult for returned migrants to retrieve land. In a setting characterised by land scarcity, it is logical that migrants who permanently cede rights to their holdings will endeavour to exact some form of compensation.

This is particularly true when, as in the above Rwandan case, land sales are already prevalent. When out-migration is common, another mechanism, involving pledges, may aid the transition from temporary to permanent transfers. If land pledges are acceptable but sales are not, pledges may eventually lead to disguised sales (Shipton, 1989). First, on tacit agreement between pledgor and pledgee, land pledges may take the form of deadlined mortgages; the transfer of land becomes permanent if the debt has not been paid within a specified period. Second, the standard stipulation that land is redeemable by offspring upon the death of the pledgor may be waived.

When a migrant leaves land in the care of kin or neighbours, two goods are exchanged: on the one hand, the protective service provided by the caretaker; on the other, the rights to manage and cultivate relinquished by the émigré. Both of these will increase in value as land increases in scarcity. When migration is understood to be temporary and short term, the protective service provided by the caretaker may be more valuable than the right to manage and cultivate that the caretaker in turn acquires, and the direction of compensation may be from town to country, for example, in the form of reciprocal cash transfers. However, if emigration is long-term or permanent, cultivation rights will increase in value relative to the protective service, and the direction of compensation may be reversed (increasing interest rates will have the same effect).

When land is effectively purchased or leased from exiting villagers, traditional prohibitions against alienation are particularly easy to subvert. First of all, the transfer of land itself has traditional antecedents in the caretaker role provided by kin or neighbours. Second, any monetary compensation given, and its attendant manifestations in the form of investment or purchase of consumer goods, will not easily be detected since the recipient is no longer present in the transparent village economy.

Increasing migration and mobility, not just for employment pur-
poses but in general, will have the effect of severing ties between land and its holders, and may thus spawn increasing trade in land.\textsuperscript{8} It might be thought that such transactions are rather infrequent and isolated events, hardly capable of generating a sustained market in land. Again, however, it is worth recalling that in many communities the restrictions on land sales cease to apply to plots that have already been sold once. Over time, therefore, the amount of land that is permanently alienable in a village where out-migration is common may become substantial. Furthermore, migration in general, in addition to making available those plots vacated by the migrants, may increase demand for land in the areas in which the migrants settle.

The implication here is that an expansion of labour markets, in addition to providing an increased demand for land transactions for reasons outlined in Section 2, will often create further demand for such transactions through a combination of the demographic consequences of labour markets and traditional entry and exit rules, as well as find a mechanism through which this demand can be channelled by way of the traditional caretaker role.

Distress sales
Distress sales of land are caused by adverse shocks, for example, drought and subsequent famine, which compel poorer farmers to alienate all or part of their land in order to survive. Desperation of this nature may be one of the most common reasons behind outright sales of indigenous land in sub-Saharan Africa (Platteau, 1996). In accordance, Troutt (1992) found that poverty was the primary reason for sales of land in various areas of Uganda, while André and Platteau (1998), in a study from Rwanda, observed that distress sales were the main reason for recent increases in land market activity.

Distress sales will primarily occur when reciprocal mechanisms, for one reason or another, have broken down. Reciprocity may, for example, have been supplanted by market exchange for a wide range of goods but without a concomitant emergence of rural financial or insurance markets. More relevantly, adverse, covariant shocks will in and of themselves often cause a more temporary breakdown in traditional risk-spreading mechanisms (Fafchamps, 1992). Recipro-

\textsuperscript{8} Increasing opportunities for mobility and communication will also serve to link demand and supply, and may also promote the dispersal of non-traditional values and belief systems.
cal mechanisms primarily serve to spread idiosyncratic risk, and in cases of widespread famine, even those household that, relatively speaking, are better off simply do not have any surplus to distribute.  

The ‘how’ of distress sales is apparently as straightforward as the ‘why’. As noted by Dalton (1962), traditional rules governing permissible transaction modes for various bundles of goods are suspended in cases of emergency (Dalton calls such suspensions ‘conversions’). Thus, a prohibition against land sales no longer applies if such sales are required for the immediate survival of the nominal owner of the land.

As noted by Bohannan and Dalton (1962), a common (though by no means universal) feature of distress sales is that land falls into the hands of outsiders, since it is considered taboo for members of the local community to take advantage of the plight of neighbours and kin to expand land holdings; if funds are available locally, these should be extended on a reciprocal basis. In such cases, new values and rules, including those related to the legitimacy of land alienation, may be ushered into the community along with the new owners of land.

Finally, there is often a close link between distress sales and off-farm income opportunities. The very problems that force people to alienate land for cash may also force them to seek paid employment in urban centres. Another link is that identified by André and Plat-teen (1998) in Rwanda. Here, differentiation in off-farm income opportunities between poor and wealthy households meant that the latter gradually acquired increasing shares of the former’s holdings, leading inexorably to a process of land concentration.

4. Policy Implications

African sales markets for land have so far not been shown to enhance efficiency (Platteau, 2000). And as Bromley (1997) notes, individual decisions in the market may lead to aggregate social outcomes that would not have been ‘voted’ for if these were known at the time. Thus, increasing trade in land is not an objective in itself, nor is bringing about conditions under which a demand for such exchange is rising. The notion, for example, of intentionally promoting conditions that lead to an increase in distress sales is plainly

---

9 Also, when shocks are localised rather than covariant, distress migration unattended by distress sales may ensue.
nonsensical. Policy measures such as imposition of property taxes require careful consideration; although such taxes will discourage possession of idle land, they may also engender distress sales unless applied only to holdings above a certain size. More generally, as Platteau (1991a) has argued, wholesale efforts to transplant the institutions and mechanisms of the free market to communities characterised by personalised exchange relationships and ethics may be counter-productive. The customary prohibitions against trade in land common throughout SSA – communal rights to non-alienation of individual holdings – may have well-founded origins related to maintenance of important community and insurance functions.

That being said, it should also be emphasised that three of the mechanisms discussed in this article – reciprocity, investment and caretakers – may spawn efficient transactions in land. As with land reform in general, it is a matter of the scale of policy intervention; rather than a massive effort to force a particular system into what is perceived as an obstinate and archaic institutional environment, the more desirable path seems to involve gradual changes of a more organic nature, where demand for increasing trade is allowed to evolve through mechanisms already present.

Different policy implications attend the different mechanisms described. Reciprocity is an indigenous transaction mode that, among other things, serves to spread idiosyncratic risk at a collective cost of sub-optimal incentives as regards surplus accumulation and investment. Making available other options for spreading idiosyncratic risk, such as crop insurance or improved storage facilities, is therefore likely to promote the gradual conversion from a reciprocal transaction mode to market exchange. In particular, the prohibition on land sales common throughout the region may be subverted through commercial land transactions disguised as reciprocity.

It is worth pointing out that a more equal distribution of income and wealth might have the same effect, at least when increased equality tends to stabilise income at above-subsistence levels and especially when idiosyncratic shocks contribute to inequality. Policy measures aimed at reducing rural inequality may thus pave the way for trade in land. The same is, of course, true of general increases in the level of wealth, since the demand for insurance will decline. Do farmers spread risk because they are poor or are they poor because they employ wasteful risk-spreading mechanisms? The causality here may be ambiguous, but, in this sense, it is also immaterial.

Paradoxically, input markets that are unpredictable to such an extent that they reinforce traditional pooling mechanisms may in fact
represent an obstacle to wider market integration (Sjaastad, 1998). More specifically, if land markets are perceived to increase risk, that may represent an obstacle to their own emergence in the context of rural Africa; this potentially points to the need for a gradual and ‘balanced growth’ of markets, including credit and insurance opportunities that may replace more traditional pooling mechanisms. Perhaps the most important implication here, however, relates to the stability of agricultural policy and its effect on farmer expectations and risk; almost regardless of the actual content of such policy with respect to taxes, provisioning, and price interventions, strategies that are stable over time are preferable to those that are not.

In SSA, the prohibition against land alienation rarely extends to improvements attached to the land. The fixed, immobile nature of most improvements, however, means that transactions in land and in improvements are interlocked. Increased investment may therefore in and of itself create a vehicle for the subversion of traditional prohibitions against alienation of land. The critical question is how to create conditions for increased rural investment activity. With the general absence of financial markets and the lack of affordable technological packages, the experience from SSA is not encouraging. It is possible that more recent developments such as microcredit schemes and emphasis on appropriate technology will prove beneficial. In any case, the conclusion here is that any policy that successfully promotes land investment also will promote commercial land transactions.

The suggestion that increased tenure security, through for example, adjudication and furnishing of title deeds, will lead to increased investment activity can be refuted on both empirical and theoretical grounds (Place and Hazell, 1993; Pinckney and Kimuyu, 1994; Migot-Adholla et al., 1991; Besley, 1995; Sjaastad and Bromley, 1997). In SSA, investment works as a trigger for both tenure security and trade in land rather than the other way around; if promotion of investments in land is successful, increased security and trade will follow.

The final mechanism considered here involves migration and the role of caretakers; farmers who use, manage, and occasionally take permanent possession of land left behind by urban migrants. Such caretakers are often relatives of the urban migrants. Nevertheless, when land is scarce, a form of payment will often take place from caretaker to migrant. This subversion mechanism is particularly difficult to expose since the seller has left the rural community. The obvious implication is that increasing employment opportunities in
urban centres should trigger increased land market activity in rural areas. Thus, any policy measures aimed at urban job creation – whether related to conditions for establishment of small enterprises or education programmes within traditional city-based professions – will also have spillover effects in terms of land market activity in rural areas.

Increasing land scarcity and rural market integration in general may, for reasons reviewed earlier in this paper, entail an increased demand also for increasing trade in land. This paper has sought to explain how, despite numerous and important constraints to increased commercial land transactions, mechanisms already present within rural African economies may provide outlets for this demand. Policy measures related to non-tenurial aspects of the rural economy, as well as the wider economy of countries within the region, may facilitate the smooth operation of these mechanisms. It is also likely that they will be more effective and less disruptive than conventional measures, such as massive land adjudication and reform procedures, and that they will permit a more demand-driven evolution of land markets.

References
Basset, T.J., 1993, ‘Introduction: The Land Question and Agricultural Transforma-
tion in Sub-Saharan Africa’, pp. 3–34 in Basset and Crummey, eds.


Sjaastad, E., 1998, Land Tenure and Land Use in Zambia, Ås: Agricultural University of Norway, Department of Forest Science, unpublished Dr. Scient. thesis.


Economic Convergence through Savings, Trade and Technology Flows: Lessons from Recent Research

Per Botolf Maurseth

Introduction

In 1960, South Korea ranked almost on a par with Bangladesh in terms of gross domestic product (GDP) per capita: GDPs per capita in both countries were less than half the world’s average. In 1990, South Korea’s income per capita was 4.8 times that of Bangladesh and about one-third higher than the world’s average; Bangladesh’s income per capita, although 40 per cent higher than in 1960, had dropped to one-third of the international average. This was the result of Korea’s economy growing at an average of 6.4 per cent annually during these three decades, while Bangladesh’s economic growth was 1.2 per cent.

These are two specific examples, but they illustrate the importance of economic growth: differences in growth rates are decisive for prosperity and misery. In recent years there has been increasing interest in economic growth and the forces determining countries’ income levels. While growth economics was stagnant both empirically and theoretically two decades ago, there is now a large and fast-growing literature on growth theory and growth empirics. An important question analysed in this literature is whether the huge inequalities in income per capita between countries will tend to disappear or widen over time. Traditional growth economics in its simplest forms predicts convergence in per capita income levels through decreasing returns to physical and human capital. The recent literature has

Note of acknowledgements. This article has benefited from comments from Leo Andreas Grünfeld, Olav Stokke and two anonymous referees. It has been prepared as a part of the project *Globalisation – convergence or agglomeration?* sponsored by the Norwegian Research Council. Correspondence: The Norwegian Institute of International Affairs, P.O. Box 8159 Dep., N-0033 Oslo. E-mail: PerB.Maurseth@nupi.no.
identified several possible mechanisms through which convergence, or the lack of it, may occur. Generally, recent theories are less optimistic on world income differences than traditional theory. A large part of the literature predicts massive divergence, while other contributions discuss the conditions under which convergence may be an outcome. It is the aim of this article to review new and earlier literature on economic growth and discuss their implications for inequality between countries. The empirical literature will be reviewed, too, with a particular emphasis on the ability of poorer countries to catch up with richer ones.

The new and old literatures on economic growth have in common that technological change is regarded as the main driving force for growth. What distinguishes new from older theories, or endogenous from exogenous growth theories, however, is that the recent literature aims at explanations of technological change itself. This expansion of theorising widens the topic of study: growth economics is not only about how economies interact, given a certain pattern of technological change, but also about the various mechanisms through which economic interaction influences technological change.

The recent wave of new growth theories and the availability of new data have spurred a large empirical literature on growth and convergence. Roughly speaking, this literature can be classified according to three different traditions. The first is a large set of studies based on cross-country growth regressions. In this literature growth rates in a set of countries, for one or many periods, are regressed on a series of variables. These studies have revealed that, as an empirical regularity, initial income tends to reduce subsequent growth rates when other variables are accounted for. This is taken as evidence of conditional convergence. The second tradition is the study of total factor productivity, the so-called growth-accounting tradition. This approach has been used particularly intensively for growth in developed countries for which better data are available. The third tradition is the study of the dynamics of income distribution between countries. Proponents of this tradition claim that neither cross-country growth regressions nor studies of total factor productivity reveal whether income inequalities between countries increase or decrease over time.

This article will first give an overview of some basic facts about world income distribution. Thereafter I will provide a guided tour of growth theory; the focus will be on what theory has to say about convergence
and how savings, international trade and technology flows may influence the results. This will serve as a backcloth to the subsequent discussion of empirical measures of convergence and the findings in existing studies. I conclude by summing up and presenting some thoughts on what has been learnt and what we need to learn more about.

1. Divergence and Convergence: Some Stylised Facts

There are very large differences in income per capita among countries in the world.¹ In 1990, the richest country in the world was 45 times richer than the poorest. This multiple had increased from 32 in the period from 1960. These are extreme cases, of course. An inequality measure that only expresses the highest as a multiple of the lowest conceals everything in between. Still, as will become clear, massive divergence in income levels is characteristic of capitalist economic development.

The topic here is convergence versus divergence in terms of per capita income internationally, not the developments in inequality between people within individual countries. Clearly, datasets of GDP per capita reveal nothing about internal inequality. Neither will this article look at the way in which population size influences inequality between people when use is made of per capita numbers.²

It should be noted that long-term development has necessarily been characterised by divergence. The richest countries in the world have been growing – though not entirely steadily – at a rate of over 1.5 per cent annually, at least since 1870, as Angus Maddison’s long-term data show (Maddison, 1995). As argued by Lant Pritchett, this has only been possible because growth rates of developed economies have systematically been higher than those in poor countries. If growth rates in poor countries had been higher than in richer ones, the level of income in the poorest countries would have been far below subsistence levels in 1870 (Pritchett, 1997). Therefore, in the long

---

¹ The data used in this article are taken from the Penn World Tables, Mark 5.6. The data are for GDP per capita in 1960 and 1990 for 104 countries. The GDP data refer to purchasing power parities at constant international prices and are therefore comparable over time and across countries. They have been used extensively in recent research, including many of the studies reviewed in this article. The data are available at http://pwt.econ.upenn.edu/

² Melchior (2001) and Melchior and Telle (2001) discuss whether inequality between persons in the world has increased or decreased during the last decades. They find that inequality may have decreased from 1960 onwards, mainly as a result of high growth rates in populous countries, in particular China.
run, capitalist development has been characterised by divergence. This is in line with the hypothesis of Simon Kuznets (1955) of an inverted U-shaped relationship between inequality and development. Development results in higher income. Development does not happen instantaneously for all members of a statistical population (in this case countries). Instead, fractions of the population are relocated from a low-income distribution to a high-income distribution. If rich and poor countries populate the world, development will, at least partially, be characterised by countries relocating from the poor group to the rich group. Therefore, the relation between development and growth will be characterised by three elements: the nature of the low-income distribution, the nature of the high-income distribution and the consequence of relocation between the two. The last element will, at least initially, tend to increase income differences.

Recent studies of growth have tended to be occupied with short time spans, in particular the post-war period. These studies reveal the same pattern of global economic growth: there is no systematic negative relation between initial levels of income and subsequent growth (see, for example, Quah, 1993, or Barro, 1997). If there is any connection between growth and initial levels of income, it is positive. This is revealed in Figure 1, which graphs growth rates in the period from 1960 to 1990 against the log of GDP per capita level in 1960 for a sample of 104 countries.

**Figure 1.**

*Convergence, Catch-up and Divergence*

*Source: The Penn World Tables, Mark 5.6.*
2. Relative Economic Performance – Theoretical Perspectives

The figure does not support the hypothesis that there is a clear connection between initial levels of income and the subsequent growth rate. If there is any relationship, it is positive. This is demonstrated by the positive sloping regression line included in the figure. The coefficient of initial income is not significant, however.

Three types of countries are shown in the figure. The triangles represent the East Asian tiger economies, which have had very high growth rates during the last three decades. The circles represent the OECD countries. Here, there seems to be a convincing impression of a negative relationship between growth and log of initial GDP. The included regression line for these countries is negatively sloping and highly significant (at a p-level below 1 per cent). This is in accordance with conditional convergence: inequality declines between countries that share important characteristics. The squares in the figure represent the rest of the countries in the world.

Growth theory for countries should therefore be able to explain (a) weak divergence between most countries in the world, (b) very high growth rates for some countries and (c) convergence between some countries that share particular characteristics (such as the OECD countries).

The development shown in Figure 1 has been the outcome of a period that has also been characterised by a dramatic increase in world trade in goods, again according to Maddison (1995), from 8 per cent of world total GDP in 1960 to almost 14 per cent in 1990. During the same period, there has also been an enormous increase both in international direct investments and cross-border financial transactions (UN, 1999 and IMF, 1997).

Growth theory and convergence: a selective review

The economic destinies of countries have long been of major interest to economists. I will review some main conclusions from both recent and older growth theories in order to highlight where they differ and how they might contribute to an understanding of the development just described.

Most theories on economic growth rely on some notion of either physical or human capital. Economies use some of their disposable

---

3 There is no consensus, however, as to whether the recent wave of globalisation has resulted in larger net capital flows compared with earlier periods. See Obstfeld (1998).
income on savings. Savings are translated into investments that result in increased capacity for production. Therefore, the relationship between savings and production and returns to capital are important determinants of long-term economic growth. This relationship constitutes one very important demarcation line in growth theory.

**Neo-classical versus endogenous growth**

The traditional neo-classical growth models that emerged in the 1950s are based on the neo-classical production function in which there are constant (or decreasing) returns to scale, substitution possibilities between all factors of production and decreasing returns to all of them individually (see, for example, Solow, 1956). Constant returns to scale is the case when a doubling of the factors of production results in an exact doubling of production. The returns to individual factors of production refer to the increase in production of an increase in the use of that factor when other factors of production are held constant. Decreasing returns to capital is therefore the situation in which a certain increase in capital results in a larger increase in production when the use of capital is initially small. Constant returns to scale and decreasing returns to each factor of production make the model consistent with decentralised markets. In the neo-classical growth models technological change is assumed to be exogenous and equal to all production entities. Solow’s model demonstrated that equilibrium growth was not a knife-edge problem of balancing growth of the labour force with growth in physical capital due to investments.

In the neo-classical growth model, the engine of growth in the short run is capital accumulation. Through savings and investments, a country increases its production capacity. Since decreasing returns to each factor of production are assumed, the incremental gain from capital decreases as production becomes more capital-intensive. The only source of increased per capita income in the long run is technological progress, meaning that more is produced with the same amount of factors of production.

This may be illustrated in terms of the most simple neo-classical growth model. Let production be according to a Cobb–Douglas function, assume a constant savings rate and disregard depreciation. Let a dot above a variable denote the derivative with respect to time. In this case the economy will be characterised by the following equations:
In the equations $Y$ denotes production, $A$ denotes the economy's technological level, $K$ is capital, $L$ is labour and $s$ is the savings rate. $\alpha$ is the share of capital in production.

The lower equation describes the growth rate in production per capita. This in turn is will be increasing in the savings rate. As capital accumulates, however, the contribution from savings will decrease. This is reflected in the term $\frac{sY}{K} - \frac{L}{L}$ which decreases with $K$. In the long run the second term in the last equation will equal zero. Therefore, in this model, technological change is the only potential source of growth in income per capita in the long run. Technological change is given by the growth rate of $A$.

For relative growth performance, the predictions of the neo-classical model are clear-cut: In the very long run, all countries will achieve the same growth rate in per capita income. In the absence of exogenous technological progress, growth will cease in the long run, and all countries will converge towards the same level of income per capita, given that they have the same savings rate. Before all countries have achieved this level of income, poorer countries are predicted to grow faster than richer ones, as poorer countries have less capital-intensive economies and enjoy higher returns on their investments.

The above predictions are questionable. First, savings rates may vary. Second, that countries' macro-production functions are Cobb–Douglas, or that production is due to decreasing returns on capital at all, are both no more than assumptions.

Figure 2 illustrates the critical role of these two assumptions. The vertical axes denote growth (in total income). The horizontal axes denote capital intensity in the economy (defined as capital per worker). In part A of the figure, the traditional neo-classical world is graphed. The downward sloping line shows the contribution from savings. As the economy grows and becomes more capital-intensive, the contribution from savings decreases. At the point where this contribution equals the growth rate of the population, growth in per capita income vanishes. If the capital intensity grows above the equilibrium level, it will fall back to this level. The dynamics are illustrated by the arrows below the graphs. The dotted line in panel A indicates the effect of reduced savings rates: the level of income per capita
decreases but the mechanism that reduces the long-run growth rate remains.

Part B illustrates the possibility that contribution from capital accumulation first falls, then rises and thereafter falls again. There might be several reasons for a pattern like this; one is that savings vary with income. Another is that as an economy grows structural changes may push it from phases of decreasing returns to phases of increasing returns. Thereafter, as the economy grows modern, it encounters diminishing returns. Part B is a graph depicting three equilibria. The first is a poverty trap. If capital intensity increases above this equilibrium, the resulting capital accumulation will be too small to sustain the implied income per capita. Therefore growth in income per capita will be negative and the economy falls back into the poverty trap. The second equilibrium is an unstable one. Slight deviations from this equilibrium will either force the economy back into the poverty trap or to the third equilibrium in which income is higher and stable.

The possibility of constant returns to capital is graphed in part C of the figure. In this case savings determine the long-term growth rate. If the contribution from savings is higher than the population growth (as illustrated in the figure), there will be constant growth in income per capita. If the contribution from savings is lower than the population growth, there will be negative growth and production goes to zero. It is important that constant returns from savings normally result in divergence. Savings determine growth rates and there is nothing that ensures similar savings rates in different countries. As I will come back to, one important contribution from recent growth theory is that it explains how, in different ways, constant returns from savings, either in physical or in human capital, can be plausible.

The neo-classical growth model describes closed economies. If a country opens its doors to international trade, it experiences a once-and-for-all income gain due to increased static efficiency. Ventura (1997) demonstrates that international trade also has dynamic effects. If international trade results in factor price equalisation, decreasing returns to capital will only apply for the world on average and not for individual countries. The reason is that capital accumulation will not increase production in all industries but only in those that are most capital-intensive (as predicted by the Rybczynski theorem). Thus, when international trade induces factor price equalisation, the traditional source of convergence disappears. However, a weak form of convergence will still be present as more and more countries become more capital-intensive.
Figure 2. Savings and growth

A

growth

B

growth

C

growth

\( \frac{\dot{L}}{L} \)

\( sY/K \)

\( k \)

\( \frac{\dot{L}}{L} \)

\( sY/K \)

\( k \)

\( sY/K \)

\( \frac{\dot{L}}{L} \)

\( K \)
Financial integration is predicted to result in fast convergence, however. If capital moves to wherever returns are the highest, poor and capital-deficient countries will receive inflows of capital. In fact, convergence is predicted to be instantaneous in the case of complete capital mobility.

Escaping decreasing returns

The hypothesis of convergence in income per capita levels is the result of the assumption of decreasing returns to accumulatable factors of production (capital above). In the long run, growth is dependent on exogenous productivity growth. Endogenous growth theories attempt to explain technological progress as an inherent part of economic mechanisms. They incorporate some of the peculiar characteristics of technology and knowledge.

First, it is taken into account that technological progress is a produced good. Within the class of endogenous growth models two different sources of knowledge creation are being analysed. The first is *deliberate* production of knowledge. Research and development result in new knowledge that is used to produce new or better goods, or to improve productivity in goods production. The second is denoted as *learning by doing*: knowledge is produced unconsciously as people learn from each other and pick up new ideas from others’ experience.

Second, it is taken into account that knowledge is a very special good in economic terms. Knowledge can be used without being exhausted. Thus, it is a so-called *non-rival good*. Knowledge is also cumulative. New knowledge is based on results obtained previously. In this sense, we are standing on ‘the shoulders of a giant’ (Caballero and Jaffe, 1993).

Third, knowledge is to a certain extent, but not completely, an exclusive good. It is, in different ways, possible to limit others’ access to newly developed knowledge, but despite secrecy and patent protection, very often it is difficult to protect property rights to knowledge for longer periods. Both the deficient exclusiveness and the cumulative aspects imply that there are externalities connected with the production of knowledge.

---

4 Barro *et al.* (1995) discuss capital mobility in neo-classical growth models. They show that if only a part of capital is internationally mobile, the rate of convergence will slow down as compared to the case when all types of capital are mobile.
Such externalities or technological spillovers form one of the foundations for endogenous growth models. In short, they provide a basis for understanding how increasing returns may be consistent with decentralised markets (see Romer, 1986, and Barro and Sala-I-Martin, 1995). When there are technological spillovers, returns to investments in human capital may be increasing for the overall economy, while decreasing for the individual economic agents. This may be illustrated by thinking of the production function above as the production function of individual firms, represented by the subscript i in the first equation below. The level of technology in society might well be a function of the capital per worker in society (K/L) (as illustrated in the second equation). In this case the model may be formulated as:

\[
Y_i = AK_i^\alpha L_i^{1-\alpha}, \quad A = \bar{A} \left( \frac{K}{L} \right)^{\delta}, \quad \alpha + \delta = 1
\]

\[
\hat{Y} - \frac{\hat{L}}{L} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} = \left( \frac{sY - \frac{\hat{L}}{L}}{\frac{\dot{K}}{K} - \frac{\dot{L}}{L}} \right) = \left( \frac{s\bar{A} - \frac{\hat{L}}{L}}{L} \right)
\]

Thus, individual firms face diminishing returns to K_i and L_i as they regard the average level of technology as exogenous. However, if all firms expand K_i, then K/L expands as well and provides a spillover that raises the productivity of all firms. In the model framework assumed here, \( \delta \) denotes the quantitative effect of this spillover effect. Here it is assumed that the capital share \( \alpha \) and the spillover parameter \( \delta \) add up to one. Therefore there are constant returns to capital at the social level. If the amount of capital is doubled production is doubled as well. This is expressed in the second set of equations. In these equations, total production is expressed as the sum of individual firms’ production. The constant social returns to capital will yield endogenous growth in the long run, as illustrated in the third set of equations. This is the situation graphed in part C of Figure 2. In the present context, K_i may be interpreted as a mixture of human and physical capital or only as human capital. In this context the savings rate is decisive, not only for the level of income per capita but also for its long-term growth rate.
Complete versus incomplete spillovers
Since spillovers form one foundation for the new growth theories, their extent and scope may be determinant for whether new growth theory produces different predictions on convergence from those of the neo-classical model. When spillovers are complete, i.e., when positive externalities from knowledge are both relevant and available for all agents independent of industrial specialisation, distance and borders, there will be convergence. In this case, the difference between the neo-classical model and endogenous growth theory is that the growth rate is explained rather than being assumed. The explained growth rate will be common to all countries and technology is still a global public good.

If spillovers are confined within distinct economies, however, growth will depend on accumulated knowledge for the economy in question (Grossman and Helpman, 1991 and 1995). This applies to countries, economic sectors or regions. If spillovers are confined within country borders, growth rates between countries will be determined by the size of each individual country. Therefore growth rates between countries will normally differ. Rivera-Batiz and Romer (1991) discuss the implications of economic integration in this context. They show that with nationally bounded technology spillovers, international trade may not increase growth rates, though static efficiency gains from trade remain. If integration increases the knowledge base used in research in each country, however, integration might well increase long-term growth rates.

Lucas (1988) and Young (1991) provide two examples of growth models in which divergence occurs because of bounded spillovers and where divergence will typically be more pronounced when countries integrate. Lucas builds on Krugman (1986) and develops the framework of dynamic comparative advantages in which spillovers are confined to industries. Countries concentrate their production in sectors where they have a (static) comparative advantage. Productivity evolves over time as a function of aggregate past production. If some industries happen to have a potential for higher productivity growth than others, countries specialised in these industries will experience higher growth rates than other countries do. This introduces the possibility of diverging economic development.

In the simplest models of endogenous growth, spillovers are thought of as an automatic effect of production or investments. In other models, research activities are introduced as a distinct economic sector (see, for instance, Romer, 1990). Researchers generate innovations that are sold monopolistically as blueprints to producers of
goods. From these blueprints particular varieties are then produced and sold in a context of monopolistic competition to consumers. It is assumed that investments in R&D will occur until expected profits equal costs. In these models there are dynamic increasing returns in the R&D sector generated by technological spillovers. In particular, it is assumed that the R&D sector employs researchers who make use of aggregated knowledge available in the economy. Their products are new blueprints, but their research also adds to society’s knowledge stock. These models do not predict convergence. Growth will be an increasing function of the workforce employed in R&D and of aggregated knowledge. There will be dynamic effects of economic integration in two different ways. First, through trade an economy gains access to a larger flow of new varieties. This generates increased consumption. Second, economic integration allows national researchers to draw on a larger knowledge base in their research. This is expected to increase their efficiency. Aghion and Howitt (1992 and 1998), Klette and Griliches (1998) and Barro and Sala-I-Martin (1995, Ch. 7) take into account uncertainty of technological change. Instead of modelling research as a deterministic process, they think of it as a stochastic process.

The R&D models formalise older ideas of Joseph Schumpeter on creative destruction. The idea is that new innovations are destructive for previous innovations since they render them obsolete. The computer industry is a good example of this process. An interesting extension by Howitt (2000) is a model in which researchers’ efficiency depends on an existing international knowledge base. In Howitt’s model some countries do not undertake R&D. The model demonstrates how a country’s position on the world income ladder may depend on the resources and subsidies it devotes to R&D. In some situations some countries will not invest at all, in which case there is no growth.

**Technology gaps**

Also inspired by Schumpeter is a less formal and more heterogeneous tradition of studies of technological change and economic growth. Such approaches stress the ability of countries that are not at the technological forefront to adapt and imitate new technologies. The ability of poorer countries to make use of technology developed elsewhere is a function not only of the rate of innovation at the technological forefront, but is also assumed to depend on

---

5 Schumpeter (1934) and (1944).
their own absorptive capacity and their technological congruence (Abramovitz, 1994). Thus, it is expected that the extent to which poorer countries make use of technology flows from more advanced countries is a function of these poorer countries’ institutions, history, social conditions, etc. Among other factors, the level of education and human capital is assumed to be a decisive factor. This is a consequence of the assumption that technology flows not only bring outdated blueprints, but are also a source of new technological development. Thus, catch-up is viewed as a process in which poorer countries both imitate and adapt older technology.

Theories of technology gaps incorporate Posner’s and Vernon’s theories on economic development (Posner, 1961; Vernon, 1966) into a Schumpeterian view on innovation and imitation. The idea is that new technology is developed in certain countries that are constantly at the technological forefront. Later on in the product cycle, production is relocated to other countries. This may be the effect of two independent factors. First, as the advanced country keeps on innovating, efficiency and wages increase. Therefore productions of some goods become unprofitable. The relocation of the Western European textile industry to low-wage countries is an example of this mechanism. Second, as a technology grows old, it becomes well known. The technology changes nature from a semi-private to a public good. As a consequence of these conditions, other countries further down on the productivity ladder take over production of the older goods. Krugman (1979) formalises the diffusion effect; in a later study (Krugman, 1986) he analyses the crowding out effect. Barro and Sala-I-Martín (1997) present a model in which technology gaps and diffusion of technology to poor countries are combined with endogenous innovation in the leading country. An interesting implication in these models is that increased productivity at the forefront is always of benefit to both rich and poor countries. Diffusion of knowledge, however, benefits poorer countries but not necessarily richer ones. Global intellectual property rights, imposed on all countries through the WTO’s Trade-Related Intellectual Property Rights agreement (TRIPS), are an example of how global politics have traded off these considerations.  

---

6 The TRIPS agreement obliges all member countries in the WTO to establish a standardised patent institution. In principle, all patentable innovations can be patented in all countries. Maskus (2000) discusses the consequences of the TRIPS agreement. Helpman (1993) extends the technology gap model of Krugman (1986) with a discussion of intellectual property rights and their implications for the ability of poor countries to catch up.
In Fagerberg (1988) growth in a set of countries is assumed to be a function of technological distance between the country in question and the world economic leader (the US) and of the resources devoted to increasing the country’s absorptive capacity. Fagerberg demonstrates that the outcome of economic development might be both convergence and divergence. He proposes that a country’s income level will depend on its own R&D, diffusion of knowledge from abroad and the country’s capacity to exploit foreign knowledge. The technology gap hypothesis is that countries lagging far behind the frontier have a larger potential for catch-up than other countries. The frontier is taken to be knowledge in the leading economy in the world. In this set-up, therefore, growth will depend on a country’s initial income (which indicates the technology gap), its absorptive capacity and possibly some other variables. The empirical implications of this model are very similar to the empirical formulations of the neo-classical growth model. In the technology gap models, poor countries are predicted to have a high potential for growth through technology imports; in neo-classical models, they are predicted to grow fast because of high returns to capital.

Verspagen (1991) models catch-up and technology flows in a similar way. He explicitly allows for the existence of underdevelopment traps. In the case of countries that are way behind the technological leader, their ability to make use of technology flows is limited. Other countries, further up the productivity ladder, have higher absorptive capacity and are able to keep the technology gap constant or reduce it. Thus, Verspagen’s model predicts a world in which there is a club of very poor countries and another club of converging wealthy countries.

**Summing up**
Recent growth theory is to a less extent than traditional theory based on assumptions of decreasing returns on physical or human capital. Leaving behind that assumption also implies that the traditional source of convergence vanishes. In a large class of models, convergence in income per capita is shown to be dependent on whether technology flows are global or local in scope and whether knowledge spills over between industries. Moreover, when there is international trade, convergence depends on the extent to which prices of goods imported from technological leaders tend to fall over time as technology progresses.
3. Empirical Evidence

Measurement and methodology
In the empirical literature several measures of convergence have been proposed. The first has already been mentioned: the lack of an unconditional systematic relationship between the initial level of GDP and subsequent growth rates for the world economy is referred to as unconditional $\beta$-divergence. Conditional $\beta$-convergence is the occurrence of convergence when other factors are controlled for. $\beta$-convergence, therefore, denotes a negative coefficient for initial level of GDP in a cross-section regression on growth rates for a sample of countries according to the regression equation:

$$g = \frac{1}{T} \log \left( \frac{y_{it}}{y_{i,t-T}} \right) = \alpha + \beta \log(y_{i,t-T}) + \gamma X_{i,t-T} + u$$

Above, $y_{it}$ denotes GDP per capita in entity $i$ at time $t$. $T$ denotes the time from the initial year to the last year; $u$ is the regression residual. The regression equation above therefore expresses the hypothesis that growth depends in the (log of) initial income and a set of other variables. One distinguishes between conditional $\beta$-convergence and unconditional $\beta$-convergence according to whether other relevant variables, denoted by the vector $X$, are included or not. Unconditional $\beta$-convergence means that $\beta$ is negative and significant when $X$ is left out. Conditional $\beta$-convergence means that $\beta$ is negative and significant when other explanatory variables are also included in the regression. The literature is not conclusive on what variables to include. Often included are variables reflecting openness to trade, the population’s educational level and the level of investments. Levine and Renelt (1992) and Barro (1997) provide critical reviews of what conditioning variables to include in cross-country growth regressions. Dobson et al. (2001) argue that the rate of convergence obtained in such regressions depends on what conditioning variables are included. In the following subsection, I give an overview of some empirical results in this tradition of cross-sectional studies of economic growth.

The reader should note that the above expression might capture the neo-classical hypothesis of convergence, the endogenous growth hypothesis with international technology diffusion and the technology gap models (when a lag to a frontier is included).

A more restrictive version of convergence is so-called $\sigma$-con-
Economic Convergence through Savings, Trade and Technology Flows

σ-convergence denotes that the standard deviation of (the log of) GDP per capita in a sample of countries decreases over time. σ-convergence is a stronger criterion than β-convergence in the sense that absence of σ-divergence can co-exist with β-convergence but not the other way around. The relation between β-convergence and σ-convergence may be derived from the above equation. Rewriting it and setting T=1, a equation of log(yₜ) is obtained. This equation is a so-called difference equation in which the level in one period depends on the level in the previous period. The term u remains in the equation and is assumed to be a random variable with zero mean and constant variance over time and over our units of observation (absence of autocorrelation and heteroscedasticity). Taking the sample variance of this expression gives:

\[ \sigma^2_{y_t} = (1 + \beta)^2 \sigma^2_{y_{t-1}} + \sigma^2_u \]

Above, \( \sigma^2_{y_t} \) denotes sample variance of the log of GDP per capita in year t and \( \sigma^2_u \) is the sample variance of u. It is seen that the expression for variance in GDP levels per capita is a function of \( \beta \). If \( \beta \) is negative (as implied by the β-convergence hypothesis), it contributes to reduced sample variance over time. Variance might nevertheless increase if the contribution from the error term, u, is larger than the contribution from β-convergence.

The second tradition of empirical studies I will review is the analyses of total factor productivity. From the production function presented earlier we have:

\[ Y = AK^{\alpha}L^{1-\alpha} \]

\[ \frac{\dot{A}}{A} = \frac{\dot{Y}}{Y} - \alpha \frac{\dot{K}}{K} - (1-\alpha) \frac{\dot{L}}{L} = \gamma X + u \]

From the second of these equations, growth in total factor productivity is expressed as the difference between growth in total levels of GDP and a weighted average of factors of production (capital and labour in this simple stylised example). This is similar to the expressions above. The last of these expresses the hypothesis that growth in total factor productivity is a linear function of possible explanatory variables.

Normally it is assumed that factors of production are paid their marginal productivity. This means that capital and workers are em-
ployed until the value of the extra production that results equals the cost of hiring them. In this case workers’ share of production is equal to \((1-\alpha)\) and capital’s share of production is equal to \(a\). Therefore, growth in productivity will be equal to the difference between the growth rates in GDP and the reward to the factors of production, times the growth rates in these. In many countries both investments and wages are observable. Given the above assumptions, total factor productivity can be estimated. Growth in factor productivity is used as a dependent variable in this kind of study. This is the growth-accounting procedure.

Studies of total factor productivity have revealed that growth in total factor productivity is substantial. In fact, several studies have demonstrated that productivity growth accounts for the major share of growth. Growth in total factor productivity has been denoted a measure of our ignorance (Abramowitz, 1956) because it is the share of growth that cannot be accounted for by growth in traditional factors of production. In recent research, however, it is very often the productivity that is subject to research.

There are important limitations to growth accounting and studies of total factor productivity. This approach is based on assumptions of constant returns to scale in production and of perfect competition. Barro and Sala-I-Martin (1995) also point out that growth in capital and production might be the consequence of growth in total factor productivity. If so, the usual measures of total factor productivity underestimate the contribution from technological change and overestimate the contribution from capital accumulation.

**Empirical results – an overview**

*Growth regressions*. Growth regressions have been very popular in recent years. There are two traditions of growth regressions on data sets for global data. The first attempts to test the neo-classical growth model, often extended with human capital. These studies indicate that the level of GDP per capita can be well explained only by inclusion of investments and human capital (see, for example, Mankiw *et al.*, 1992). However, these two variables do not succeed in explaining growth, i.e., changes in levels, very well.

The second approach to global data sets has been to include a large set of explanatory variables in regressions on growth. These exercises have been useful in at least two senses. First, they reveal possible explanations for growth. In growth regressions, investments, schooling (male, but not female!) and initial income are robust var-
variables correlating with growth in many studies. 7 Resource abundance is negatively related to growth, at least in countries with institutions of low quality (Sachs and Warner, 1995b and Mehlum et al., 2002). 8 Political instability is detected as important for economic growth, although it is hard to determine whether this reflects the impact of social unrest, insecure property rights or lack of other institutional qualities. Openness to trade is a less robust explanatory variable, but several studies indicate a strong, but not very significant effect. Rodriguez and Rodrik (1999) present evidence of the opposite: they find no significant relationship between trade policy and economic growth. Second, growth regressions clarify the concept of convergence: by use of such regressions on different samples of countries and with different explanatory variables, one may detect to what extent initial income robustly influences subsequent growth.

Growth regressions are not without problems. I will emphasise three of them. First, it is not clear what the direction of causation between the explanatory variables and growth is. Neither is it clear what variables to include in growth regressions. Levine and Renelt (1992) have constructed a test for the robustness of explanatory variables in growth regression. The essence of their test is that a variable should be statistically significant and of the same sign in regressions independently of the inclusion of other different variables. This implies that investment will be a robust explanatory variable if regressions give a positive and significant result independently of whether other variables, like schooling, trade policy etc. are included. Second, growth regressions of the type cited below very often presume that countries can be observed independently. The most common regression methods are based on ordinary least squares regression and it is not taken into account how countries interact with each other. Third, growth regressions have limited explanatory power. One reason for this is that regressions on the largest samples possible provide researchers with a small set of available explanatory variables. We believe that investments in R&D are an important source of growth, but for many countries R&D data are not available. Investments in human capital are therefore often approximated, for instance, by data on school enrolment.

---

7 The finding that female schooling is associated with low growth rates has spurred debate and further studies. There is now agreement that it should not to be interpreted as a causal effect, but it represents a puzzle in the data. See Klasen (2002).

8 It is found in several studies that a high share of exports of natural resources in a country’s GDP is negatively correlated with economic growth. Two main explanations have been put forward. The first is Dutch disease, which implies that other productive activities are crowded out. The second is that resource abundance stimulates rent-seeking activities.
Table 1 reports results from some important studies. It is seen that only three variables stand out as robust explanatory variables of growth. These are initial income, investments and international trade. Other variables are often not significant or their significance (and even their sign) depends on what other variables are included. Often variables seem to have non-linear effects. This is the case both for indexes of democracy and for inequality.

### Studies of total factor productivity

In empirical studies of factor productivity convergence is not the issue. The focus is on productivity and its determinants. In studies like these, a hypothesis that is often tested is the predicted potential for lagging countries,

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reference</th>
<th>Effect</th>
<th>Robust/Fragile</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) initial income</td>
<td>Ba, MK, B, BS, I, SW, M</td>
<td>-*</td>
<td>R</td>
</tr>
<tr>
<td>(2) investments</td>
<td>Ba, MK, B, BS, I, S</td>
<td>+*</td>
<td>R</td>
</tr>
<tr>
<td>(3) human capital</td>
<td>Ba, MK, B, I</td>
<td>+ (- )*</td>
<td></td>
</tr>
<tr>
<td>(4) trade</td>
<td>FR</td>
<td>+*</td>
<td>R</td>
</tr>
<tr>
<td>(5) trade policy</td>
<td>S, SW (1), RR</td>
<td>+(-)*</td>
<td>F</td>
</tr>
<tr>
<td>(6) foreign direct investments</td>
<td>BLZ</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>(7) corruption</td>
<td>MK</td>
<td>-*</td>
<td></td>
</tr>
<tr>
<td>(8) democracy</td>
<td>B</td>
<td>±*</td>
<td></td>
</tr>
<tr>
<td>(10) health</td>
<td>B</td>
<td>+*</td>
<td></td>
</tr>
<tr>
<td>(11) inequality</td>
<td>B(1), PT</td>
<td>±*</td>
<td></td>
</tr>
<tr>
<td>(12) inflation</td>
<td>LR</td>
<td>-</td>
<td>F</td>
</tr>
<tr>
<td>(13) regions</td>
<td>SM, B</td>
<td>+* (East Asia,lat)</td>
<td></td>
</tr>
<tr>
<td>(14) rule of law</td>
<td>SM, B</td>
<td>+*</td>
<td></td>
</tr>
<tr>
<td>(15) religion</td>
<td>SM</td>
<td>-(±) Christianity</td>
<td></td>
</tr>
<tr>
<td>(16) size of public sector</td>
<td>Ba, BS</td>
<td>-*</td>
<td></td>
</tr>
<tr>
<td>(17) resources</td>
<td>SW, M</td>
<td>-(±)</td>
<td></td>
</tr>
<tr>
<td>(18) political stability</td>
<td>SW</td>
<td>+*</td>
<td></td>
</tr>
</tbody>
</table>

sectors or firms to catch up in terms of productivity by use of knowledge developed elsewhere. In order to study the effects of innovation and knowledge flows or spillovers, as modelled in endogenous growth models, many researchers have chosen to focus on smaller data sets for which more variables are available. Such variables are data on R&D, patents and, most important for our subjects, those reflecting the diffusion of technology. I will distinguish between findings of embodied and disembodied technology flows since their interpretations differ.

As discussed above, technology flows potentially have many forms. One is technology flows embodied in goods. Buyers benefit from the knowledge that is used to develop a good, both if the good is used as a factor in production and if it is used for consumption. A set of studies has revealed important effects of embodied technology flows for growth in factor productivity.

(A) Coe and Helpman (1995) hypothesise that growth in productivity depends positively on a country’s own R&D and other countries’ R&D. They assume that others’ R&D is imported through imports of capital goods. They therefore regress productivity growth in the OECD countries on each country’s own R&D and a weighted sum of other countries’ R&D where the weights are the shares of imports from those countries to the country in question. The results are striking: Coe and Helpman find that most productivity growth results from foreign R&D and not from national R&D. The import of foreign R&D has greater influence on smaller countries than on large ones. A later study is that of Frantzen (2001), who extended Coe and Helpman’s analysis to a longer period.

(B) Coe, Helpman and Hoffmeister (1997) extend the above study to a group of developing countries. In this study, evidence is found that foreign R&D stocks and imports of capital goods from other countries explain growth in total factor productivity more than does, for instance, schooling. Furthermore, the effect of foreign R&D seems to be larger the more open the economy is.

(C) Lichtenberg and van Pottelsberghe de la Potterie (1996) aim at extending the analysis by Coe and Helpman to flows of international foreign direct investments. Their findings do not lend support to the view that important technology flows from the investing country to the recipient country. Their findings suggest the opposite; the investing country benefits from R&D in the
host country. This is an important finding since many proponents of foreign direct investments (FDI) argue that inward FDI are an important source of technology. In fact, several studies indicate that this is not the case.

(D) Another extension of Coe and Helpman is the study by Eaton and Kortum (2001). They hypothesise that imports of capital goods depend on these goods’ prices. They estimate a price index of trade capital goods for importing countries and find that countries that face high prices of imported capital goods experience lower productivity and income per capita.

(E) Similar results are found in Maurseth (2003) in which a theoretical price index of capital goods is constructed. The price index is constructed according to an assumption that geographical distance is an important barrier to trade. Therefore, the price index of capital will be higher in peripheral countries. This gives an explanation for the empirical regularity that high-income countries are located near large markets. Similarly these results indicate that growth rates between countries will be geographically clustered. Negative growth in one country infects neighbouring countries and positive growth is likewise transmitted to neighbours.

(F) These results are in line with Easterly and Levine (1998) who explicitly estimate contagion effects in economic growth. They find that countries are affected by the growth destiny of their neighbours. For the world economy, the nature of these findings contribute to explanations of why clusters of countries get rich and other clusters remain poor.

There is also another set of studies that focuses on disembodied technology flows. These denote flows of technology that occurs without economic transactions as prerequisites. Examples of such technology flows are exchange of knowledge in academic research, industrial espionage and reverse engineering.

(A) In models of technology gaps, the main hypothesis is that a technology gap between a poor country and the leading country potentially favours growth in the poor country. Fagerberg (1987) demonstrates that, for a sample of 25 countries, including the OECD countries, growth is well explained as a positive function
of each country’s number of patents (as a measure of innovation), a negative function of the technology level (measured as the country’s own GDP per capita) and investments. The negative coefficient of initial level of GDP is interpreted as a technology gap between the country in question and the technology leader in the sample (the US). It should be noted that this study does not differ from growth regressions except for the inclusion of patents as an indicator of technology. The interpretations of the result differ, however.

(B) In the same vein, Griffith, Redding and Van Reenen (2000) estimate productivity in industries in a country as a function of the lag between productivity of the industry in this country and the productivity of the same industry in the country with the highest productivity in that industry. They find clear evidence of convergence in productivity levels between countries.

(C) Eaton and Kortum (1996) analyse international patenting. They hypothesise that if an invention is patented in a country (particularly when it is not where the invention originated), it signals a transfer of technology. They estimate the determinants of international patenting and find, among other things, that distance reduces knowledge diffusion. They find positive and significant effects of international knowledge flows in the same vein as Coe and Helpman (1995): foreign innovation is more important than national innovation in smaller countries. Eaton and Kortum analyse growth in labour productivity, however.

(D) Keller (2002) estimates total factor productivity as a function of a country’s own R&D and that of others, in 14 countries, but for a large set of industries. He finds that the effects of others’ R&D on a sector’s productivity decrease rapidly with geographical distance and linguistic borders.

(E) Verspagen (1997) estimates total factor productivity in different industries and uses patent citations as the weights for technology diffusion from one sector to another. Verspagen’s analysis, there also seem to be important effects of technology diffusion. The same result is found in Maurseth (2001) for a disaggregated set of Western European regions.

To sum up: studies of total factor productivity suggest that the pro-
ductivity in industries and countries depends to a large degree on technology flows from other sources rather than from their own inventions.

\textbf{σ-convergence and other types of distribution dynamics.} As mentioned above, a strict test of convergence is σ-convergence. σ-convergence denotes reduced standard deviation in the cross-country income distribution over time. As such the measure is extremely simple. There have been only a few studies that incorporate explanatory variables in analyses of σ-convergence. Two of these are Ben-David (1996) and Ben-David and Kimhy (2001). Ben-David acknowledges the problems of including trade in growth regressions. He therefore analyses σ-convergence among trading partners. In particular, he finds that pairs of countries that trade intensively with each other show less dispersion in their income than other countries. Similarly, he finds that pairs of countries that increase their trade relations, experience reduced dispersion in their income per capita.

A related finding is presented in Figure 3. The figure shows the dispersion in income per capita among countries standardized to world average and in income per capita standardized to a distance-weighted world average. In analyses of geography in general (and for economic growth in particular), the hypothesis is that some variable x in entity i influences some variable y in entity j as a decreasing function of the distance from i to j, $d_{ij}$. Therefore, a distance weights matrix was constructed according to:

$$w_{ij} = \frac{1/d_{ij}}{\sum_{j=1}^{n} 1/d_{ij}}$$

The resulting weight matrix postulates that the influence of any variable between two countries decreases with the inverse of the distance between them. The weights are standardized so that they add up to one for each country. This makes it easier to construct weighted averages of variables for countries.

The figure reveals that dispersion is less between neighbours but that s-divergence occurs in both the overall distribution and the distance-normalised distribution.
Quah (1993 and 1996) argues that both $\beta$-convergence and $\sigma$-convergence are crude measures of convergence. For instance, both can be consistent with Baumol’s notion of convergence clubs, in which there are clubs of countries converging towards common levels of GDP per capita (Baumol, 1986). Quah proposes to report transition probabilities from percentiles of the distribution of income over time. Thus, growth clubs would be characterised by more entries into certain percentiles of the population than exits from the same percentiles. The essence of his proposal is demonstrated in Figure 4, which graphs the ranking of 104 countries in the world economy in 1960 and 1990. Quah’s transition probabilities correspond to countries jumping from one of the graphed squares to another. He characterises the cross-country income distribution as stable if countries remain within those squares and unstable if they jump out of their squares. An objection to this approach is that countries at the lower and higher parts of the income distribution can only ‘jump’ in one direction. Still, comparing the ranking of countries at different points in time gives an impression of stability versus instability.
income distribution across countries in the world was more stable for rich countries than for others. This reflects the clear convergence among the rich OECD countries.

4. Summary and Conclusion
Whether countries will tend to converge in income per capita is an important question for students of economic growth. While convergence was an inherent prediction in the traditional neo-classical growth model because of decreasing returns to capital, in recent theories convergence is predicted to depend on diffusion of knowledge. Diffusion of knowledge takes many forms and is often distinguished as being embodied in traded goods and disembodied flows of knowledge.

Recent empirical research lends support to the neo-classical hypothesis of conditional convergence: when other relevant factors are accounted for, there is convergence in GDP per capita. It is not clear from growth regressions what to conclude from this. One
interpretation is that this supports the hypothesis of decreasing returns to capital. Another is that low levels of initial income indicate a large potential for catch-up through assimilation of technology. It is important that conditional convergence is not equivalent with a collapsing or narrowing income distribution. In fact, differences between rich and poor countries have increased. Growth regressions have revealed important potential sources of growth, however. These are investments in human and physical capital, institutional quality and openness to trade.

Studies of smaller datasets demonstrate a potentially large influence of technology diffusion. Of the channels for knowledge spillovers, trade between countries has been identified as important. It is not clear from recent studies whether trade-induced spillovers dominate in importance over the disembodied spillovers analysed in the first generations of endogenous growth models.

While growth economics has revealed a set of important mechanisms related to economic growth, it has not resulted in a toolbox for growth-promoting policies. In particular for very poor countries, there are many remaining questions. The effects of institutional quality, governance and geography on economic growth seem to be major issues for future research.

References


Penn World Tables, Mark 5.6 is available at http://pwt.econ.upenn.edu/


Economic Convergence through Savings, Trade and Technology Flows: Lessons from Recent Research

Per Botolf Maurseth

Introduction

In 1960, South Korea ranked almost on a par with Bangladesh in terms of gross domestic product (GDP) per capita: GDPs per capita in both countries were less than half the world’s average. In 1990, South Korea’s income per capita was 4.8 times that of Bangladesh and about one-third higher than the world’s average; Bangladesh’s income per capita, although 40 per cent higher than in 1960, had dropped to one-third of the international average. This was the result of Korea’s economy growing at an average of 6.4 per cent annually during these three decades, while Bangladesh’s economic growth was 1.2 per cent.

These are two specific examples, but they illustrate the importance of economic growth: differences in growth rates are decisive for prosperity and misery. In recent years there has been increasing interest in economic growth and the forces determining countries’ income levels. While growth economics was stagnant both empirically and theoretically two decades ago, there is now a large and fast-growing literature on growth theory and growth empirics. An important question analysed in this literature is whether the huge inequalities in income per capita between countries will tend to disappear or widen over time. Traditional growth economics in its simplest forms predicts convergence in per capita income levels through decreasing returns to physical and human capital. The recent literature has

Note of acknowledgements: This article has benefited from comments from Leo Andreas Grünfeld, Olav Stokke and two anonymous referees. It has been prepared as a part of the project Globalisation – convergence or agglomeration? sponsored by the Norwegian Research Council. Correspondence: The Norwegian Institute of International Affairs, P.O. Box 8159 Dep., N-0033 Oslo. E-mail: PerB.Maurseth@nupi.no.
identified several possible mechanisms through which convergence, or the lack of it, may occur. Generally, recent theories are less optimistic on world income differences than traditional theory. A large part of the literature predicts massive divergence, while other contributions discuss the conditions under which convergence may be an outcome. It is the aim of this article to review new and earlier literature on economic growth and discuss their implications for inequality between countries. The empirical literature will be reviewed, too, with a particular emphasis on the ability of poorer countries to catch up with richer ones.

The new and old literatures on economic growth have in common that technological change is regarded as the main driving force for growth. What distinguishes new from older theories, or endogenous from exogenous growth theories, however, is that the recent literature aims at explanations of technological change itself. This expansion of theorising widens the topic of study: growth economics is not only about how economies interact, given a certain pattern of technological change, but also about the various mechanisms through which economic interaction influences technological change.

The recent wave of new growth theories and the availability of new data have spurred a large empirical literature on growth and convergence. Roughly speaking, this literature can be classified according to three different traditions. The first is a large set of studies based on cross-country growth regressions. In this literature growth rates in a set of countries, for one or many periods, are regressed on a series of variables. These studies have revealed that, as an empirical regularity, initial income tends to reduce subsequent growth rates when other variables are accounted for. This is taken as evidence of conditional convergence. The second tradition is the study of total factor productivity, the so-called growth-accounting tradition. This approach relies more on stringent (and controversial) theoretical assumptions but has the potential to explain determinants not only of income, but also of productivity. This approach has been used particularly intensively for growth in developed countries for which better data are available. The third tradition is the study of the dynamics of income distribution between countries. Proponents of this tradition claim that neither cross-country growth regressions nor studies of total factor productivity reveal whether income inequalities between countries increase or decrease over time.

This article will first give an overview of some basic facts about world income distribution. Thereafter I will provide a guided tour of growth theory; the focus will be on what theory has to say about convergence
and how savings, international trade and technology flows may influence the results. This will serve as a backcloth to the subsequent discussion of empirical measures of convergence and the findings in existing studies. I conclude by summing up and presenting some thoughts on what has been learnt and what we need to learn more about.

1. Divergence and Convergence: Some Stylised Facts

There are very large differences in income per capita among countries in the world. In 1990, the richest country in the world was 45 times richer than the poorest. This multiple had increased from 32 in the period from 1960. These are extreme cases, of course. An inequality measure that only expresses the highest as a multiple of the lowest conceals everything in between. Still, as will become clear, massive divergence in income levels is characteristic of capitalist economic development.

The topic here is convergence versus divergence in terms of per capita income internationally, not the developments in inequality between people within individual countries. Clearly, datasets of GDP per capita reveal nothing about internal inequality. Neither will this article look at the way in which population size influences inequality between people when use is made of per capita numbers.

It should be noted that long-term development has necessarily been characterised by divergence. The richest countries in the world have been growing – though not entirely steadily – at a rate of over 1.5 per cent annually, at least since 1870, as Angus Maddison’s long-term data show (Maddison, 1995). As argued by Lant Pritchett, this has only been possible because growth rates of developed economies have systematically been higher than those in poor countries. If growth rates in poor countries had been higher than in richer ones, the level of income in the poorest countries would have been far below subsistence levels in 1870 (Pritchett, 1997). Therefore, in the long
run, capitalist development has been characterised by divergence. This is in line with the hypothesis of Simon Kuznets (1955) of an inverted U-shaped relationship between inequality and development. Development results in higher income. Development does not happen instantaneously for all members of a statistical population (in this case countries). Instead fractions of the population are relocated from a low-income distribution to a high-income distribution. If rich and poor countries populate the world, development will, at least partially, be characterised by countries relocating from the poor group to the rich group. Therefore, the relation between development and growth will be characterised by three elements: the nature of the low-income distribution, the nature of the high-income distribution and the consequence of relocation between the two. The last element will, at least initially, tend to increase income differences.

Recent studies of growth have tended to be occupied with short time spans, in particular the post-war period. These studies reveal the same pattern of global economic growth: there is no systematic negative relation between initial levels of income and subsequent growth (see, for example, Quah, 1993, or Barro, 1997). If there is any connection between growth and initial levels of income, it is positive. This is revealed in Figure 1, which graphs growth rates in the period from 1960 to 1990 against the log of GDP per capita level in 1960 for a sample of 104 countries.

![Figure 1. Convergence, Catch-up and Divergence](source: The Penn World Tables, Mark 5.6)
2. Relative Economic Performance – Theoretical Perspectives

The figure does not support the hypothesis that there is a clear connection between initial levels of income and the subsequent growth rate. If there is any relationship, it is positive. This is demonstrated by the positive sloping regression line included in the figure. The coefficient of initial income is not significant, however.

Three types of countries are shown in the figure. The triangles represent the East Asian tiger economies, which have had very high growth rates during the last three decades. The circles represent the OECD countries. Here, there seems to be a convincing impression of a negative relationship between growth and log of initial GDP. The included regression line for these countries is negatively sloping and highly significant (at a p-level below 1 per cent). This is in accordance with conditional convergence: inequality declines between countries that share important characteristics. The squares in the figure represent the rest of the countries in the world.

Growth theory for countries should therefore be able to explain (a) weak divergence between most countries in the world, (b) very high growth rates for some countries and (c) convergence between some countries that share particular characteristics (such as the OECD countries).

The development shown in Figure 1 has been the outcome of a period that has also been characterised by a dramatic increase in world trade in goods, again according to Maddison (1995), from 8 per cent of world total GDP in 1960 to almost 14 per cent in 1990. During the same period, there has also been an enormous increase both in international direct investments and cross-border financial transactions (UN, 1999 and IMF, 1997).

Growth theory and convergence: a selective review

The economic destinies of countries have long been of major interest to economists. I will review some main conclusions from both recent and older growth theories in order to highlight where they differ and how they might contribute to an understanding of the development just described.

Most theories on economic growth rely on some notion of either physical or human capital. Economies use some of their disposable

---

3 There is no consensus, however, as to whether the recent wave of globalisation has resulted in larger net capital flows compared with earlier periods. See Obstfeld (1998).
income on savings. Savings are translated into investments that result in increased capacity for production. Therefore, the relationship between savings and production and returns to capital are important determinants of long-term economic growth. This relationship constitutes one very important demarcation line in growth theory.

**Neo-classical versus endogenous growth**

The traditional neo-classical growth models that emerged in the 1950s are based on the neo-classical production function in which there are constant (or decreasing) returns to scale, substitution possibilities between all factors of production and decreasing returns to all of them individually (see, for example, Solow, 1956). Constant returns to scale is the case when a doubling of the factors of production results in an exact doubling of production. The returns to individual factors of production refer to the increase in production of an increase in the use of that factor when other factors of production are held constant. Decreasing returns to capital is therefore the situation in which a certain increase in capital results in a larger increase in production when the use of capital is initially small. Constant returns to scale and decreasing returns to each factor of production make the model consistent with decentralised markets. In the neo-classical growth models technological change is assumed to be exogenous and equal to all production entities. Solow’s model demonstrated that equilibrium growth was not a knife-edge problem of balancing growth of the labour force with growth in physical capital due to investments.

In the neo-classical growth model, the engine of growth in the short run is capital accumulation. Through savings and investments, a country increases its production capacity. Since decreasing returns to each factor of production are assumed, the incremental gain from capital decreases as production becomes more capital-intensive. The only source of increased per capita income in the long run is technological progress, meaning that more is produced with the same amount of factors of production.

This may be illustrated in terms of the most simple neo-classical growth model. Let production be according to a Cobb–Douglas function, assume a constant savings rate and disregard depreciation. Let a dot above a variable denote the derivative with respect to time. In this case the economy will be characterised by the following equations:
In the equations $Y$ denotes production, $A$ denotes the economy’s technological level, $K$ is capital, $L$ is labour and $s$ is the savings rate. $\alpha$ is the share of capital in production.

The lower equation describes the growth rate in production per capita. This in turn is will be increasing in the savings rate. As capital accumulates, however, the contribution from savings will decrease. This is reflected in the term $(L/K)^{1-\alpha}$ which decreases with $K$. In the long run the second term in the last equation will equal zero. Therefore, in this model, technological change is the only potential source of growth in income per capita in the long run. Technological change is given by the growth rate of $A$.

For relative growth performance, the predictions of the neo-classical model are clear-cut: In the very long run, all countries will achieve the same growth rate in per capita income. In the absence of exogenous technological progress, growth will cease in the long run, and all countries will converge towards the same level of income per capita, given that they have the same savings rate. Before all countries have achieved this level of income, poorer countries are predicted to grow faster than richer ones, as poorer countries have less capital-intensive economies and enjoy higher returns on their investments.

The above predictions are questionable. First, savings rates may vary. Second, that countries’ macro-production functions are Cobb–Douglas, or that production is due to decreasing returns on capital at all, are both no more than assumptions.

Figure 2 illustrates the critical role of these two assumptions. The vertical axes denote growth (in total income). The horizontal axes denote capital intensity in the economy (defined as capital per worker). In part A of the figure, the traditional neo-classical world is graphed. The downward sloping line shows the contribution from savings. As the economy grows and becomes more capital-intensive, the contribution from savings decreases. At the point where this contribution equals the growth rate of the population, growth in per capita income vanishes. If the capital intensity grows above the equilibrium level, it will fall back to this level. The dynamics are illustrated by the arrows below the graphs. The dotted line in panel A indicates the effect of reduced savings rates: the level of income per capita
decreases but the mechanism that reduces the long-run growth rate remains.

Part B illustrates the possibility that contribution from capital accumulation first falls, then rises and thereafter falls again. There might be several reasons for a pattern like this; one is that savings vary with income. Another is that as an economy grows structural changes may push it from phases of decreasing returns to phases of increasing returns. Thereafter, as the economy grows modern, it encounters diminishing returns. Part B is a graph depicting three equilibria. The first is a poverty trap. If capital intensity increases above this equilibrium, the resulting capital accumulation will be too small to sustain the implied income per capita. Therefore growth in income per capita will be negative and the economy falls back into the poverty trap. The second equilibrium is an unstable one. Slight deviations from this equilibrium will either force the economy back into the poverty trap or to the third equilibrium in which income is higher and stable.

The possibility of constant returns to capital is graphed in part C of the figure. In this case savings determine the long-term growth rate. If the contribution from savings is higher than the population growth (as illustrated in the figure), there will be constant growth in income per capita. If the contribution from savings is lower than the population growth, there will be negative growth and production goes to zero. It is important that constant returns from savings normally result in divergence. Savings determine growth rates and there is nothing that ensures similar savings rates in different countries. As I will come back to, one important contribution from recent growth theory is that it explains how, in different ways, constant returns from savings, either in physical or in human capital, can be plausible.

The neo-classical growth model describes closed economies. If a country opens its doors to international trade, it experiences a once-and-for-all income gain due to increased static efficiency. Ventura (1997) demonstrates that international trade also has dynamic effects. If international trade results in factor price equalisation, decreasing returns to capital will only apply for the world on average and not for individual countries. The reason is that capital accumulation will not increase production in all industries but only in those that are most capital-intensive (as predicted by the Rybczynski theorem). Thus, when international trade induces factor price equalisation, the traditional source of convergence disappears. However, a weak form of convergence will still be present as more and more countries become more capital-intensive.
Figure 2. Savings and growth

A
growth

B
growth

C
growth

\( \frac{\dot{L}}{L} \)

\( sY/K \)

\( k \)

\( \frac{\dot{L}}{L} \)

\( sY/K \)

\( k \)

\( \frac{\dot{Y}}{Y} \)

\( K \)
Financial integration is predicted to result in fast convergence, however. If capital moves to wherever returns are the highest, poor and capital-deficient countries will receive inflows of capital. In fact, convergence is predicted to be instantaneous in the case of complete capital mobility.\(^4\)

**Escaping decreasing returns**

The hypothesis of convergence in income per capita levels is the result of the assumption of decreasing returns to accumulatable factors of production (capital above). In the long run, growth is dependent on exogenous productivity growth. Endogenous growth theories attempt to explain technological progress as an inherent part of economic mechanisms. They incorporate some of the peculiar characteristics of technology and knowledge.

First, it is taken into account that technological progress is a produced good. Within the class of endogenous growth models two different sources of knowledge creation are being analysed. The first is *deliberate* production of knowledge. Research and development result in new knowledge that is used to produce new or better goods, or to improve productivity in goods production. The second is denoted as *learning by doing*: knowledge is produced unconsciously as people learn from each other and pick up new ideas from others’ experience.

Second, it is taken into account that knowledge is a very special good in economic terms. Knowledge can be used without being exhausted. Thus, it is a so-called *non-rival good*. Knowledge is also cumulative. New knowledge is based on results obtained previously. In this sense, we are standing on ‘the shoulders of a giant’ (Caballero and Jaffe, 1993).

Third, knowledge is to a certain extent, but not completely, an exclusive good. It is, in different ways, possible to limit others’ access to newly developed knowledge, but despite secrecy and patent protection, very often it is difficult to protect property rights to knowledge for longer periods. Both the deficient exclusiveness and the cumulative aspects imply that there are externalities connected with the production of knowledge.

---

\(^4\) Barro *et al.* (1995) discuss capital mobility in neo-classical growth models. They show that if only a part of capital is internationally mobile, the rate of convergence will slow down as compared to the case when all types of capital are mobile.
Such externalities or technological spillovers form one of the foundations for endogenous growth models. In short, they provide a basis for understanding how increasing returns may be consistent with decentralised markets (see Romer, 1986, and Barro and Sala-I-Martin, 1995). When there are technological spillovers, returns to investments in human capital may be increasing for the overall economy, while decreasing for the individual economic agents. This may be illustrated by thinking of the production function above as the production function of individual firms, represented by the subscript i in the first equation below. The level of technology in society might well be a function of the capital per worker in society (K/L) (as illustrated in the second equation). In this case the model may be formulated as:

\[ Y_i = AK_i^\alpha L_i^{1-\alpha}, \quad A = \bar{A}\left(\frac{K}{L}\right)^\delta, \quad \alpha + \delta = 1 \]

\[ Y = \sum AK_i^\alpha L_i^{1-\alpha} = \bar{A}K^{\alpha+\delta} L^{1-\alpha-\delta} = \bar{A}K \]

\[ \frac{\dot{Y}}{Y} - \frac{\dot{L}}{L} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} = \left(\frac{sY}{K} - \frac{s\dot{L}}{L}\right) = \left(s\bar{A} - \frac{\dot{L}}{L}\right) \]

Thus, individual firms face diminishing returns to \( K_i \) and \( L_i \) as they regard the average level of technology as exogenous. However, if all firms expand \( K_i \), then \( K/L \) expands as well and provides a spillover that raises the productivity of all firms. In the model framework assumed here, \( \delta \) denotes the quantitative effect of this spillover effect. Here it is assumed that the capital share \( \alpha \) and the spillover parameter \( \delta \) add up to one. Therefore there are constant returns to capital at the social level. If the amount of capital is doubled production is doubled as well. This is expressed in the second set of equations. In these equations, total production is expressed as the sum of individual firms’ production. The constant social returns to capital will yield endogenous growth in the long run, as illustrated in the third set of equations. This is the situation graphed in part C of Figure 2. In the present context, \( K \) may be interpreted as a mixture of human and physical capital or only as human capital. In this context the savings rate is decisive, not only for the level of income per capita but also for its long-term growth rate.
Complete versus incomplete spillovers

Since spillovers form one foundation for the new growth theories, their extent and scope may be determinant for whether new growth theory produces different predictions on convergence from those of the neo-classical model. When spillovers are complete, i.e., when positive externalities from knowledge are both relevant and available for all agents independent of industrial specialisation, distance and borders, there will be convergence. In this case, the difference between the neo-classical model and endogenous growth theory is that the growth rate is explained rather than being assumed. The explained growth rate will be common to all countries and technology is still a global public good.

If spillovers are confined within distinct economies, however, growth will depend on accumulated knowledge for the economy in question (Grossman and Helpman, 1991 and 1995). This applies to countries, economic sectors or regions. If spillovers are confined within country borders, growth rates between countries will be determined by the size of each individual country. Therefore growth rates between countries will normally differ. Rivera-Batiz and Romer (1991) discuss the implications of economic integration in this context. They show that with nationally bounded technology spillovers, international trade may not increase growth rates, though static efficiency gains from trade remain. If integration increases the knowledge base used in research in each country, however, integration might well increase long-term growth rates.

Lucas (1988) and Young (1991) provide two examples of growth models in which divergence occurs because of bounded spillovers and where divergence will typically be more pronounced when countries integrate. Lucas builds on Krugman (1986) and develops the framework of dynamic comparative advantages in which spillovers are confined to industries. Countries concentrate their production in sectors where they have a (static) comparative advantage. Productivity evolves over time as a function of aggregate past production. If some industries happen to have a potential for higher productivity growth than others, countries specialised in these industries will experience higher growth rates than other countries do. This introduces the possibility of diverging economic development.

In the simplest models of endogenous growth, spillovers are thought of as an automatic effect of production or investments. In other models, research activities are introduced as a distinct economic sector (see, for instance, Romer, 1990). Researchers generate innovations that are sold monopolistically as blueprints to producers of
goods. From these blueprints particular varieties are then produced and sold in a context of monopolistic competition to consumers. It is assumed that investments in R&D will occur until expected profits equal costs. In these models there are dynamic increasing returns in the R&D sector generated by technological spillovers. In particular, it is assumed that the R&D sector employs researchers who make use of aggregated knowledge available in the economy. Their products are new blueprints, but their research also adds to society’s knowledge stock. These models do not predict convergence. Growth will be an increasing function of the workforce employed in R&D and of aggregated knowledge. There will be dynamic effects of economic integration in two different ways. First, through trade an economy gains access to a larger flow of new varieties. This generates increased consumption. Second, economic integration allows national researchers to draw on a larger knowledge base in their research. This is expected to increase their efficiency. Aghion and Howitt (1992 and 1998), Klette and Griliches (1998) and Barro and Sala-I-Martin (1995, Ch. 7) take into account uncertainty of technological change. Instead of modelling research as a deterministic process, they think of it as a stochastic process.

The R&D models formalise older ideas of Joseph Schumpeter on creative destruction. The idea is that new innovations are destructive for previous innovations since they render them obsolete. The computer industry is a good example of this process. An interesting extension by Howitt (2000) is a model in which researchers’ efficiency depends on an existing international knowledge base. In Howitt’s model some countries do not undertake R&D. The model demonstrates how a country’s position on the world income ladder may depend on the resources and subsidies it devotes to R&D. In some situations some countries will not invest at all, in which case there is no growth.

**Technology gaps**

Also inspired by Schumpeter is a less formal and more heterogeneous tradition of studies of technological change and economic growth. Such approaches stress the ability of countries that are not at the technological forefront to adapt and imitate new technologies. The ability of poorer countries to make use of technology developed elsewhere is a function not only of the rate of innovation at the technological forefront, but is also assumed to depend on

---

5 Schumpeter (1934) and (1944).
their own absorptive capacity and their technological congruence (Abramovitz, 1994). Thus, it is expected that the extent to which poorer countries make use of technology flows from more advanced countries is a function of these poorer countries’ institutions, history, social conditions, etc. Among other factors, the level of education and human capital is assumed to be a decisive factor. This is a consequence of the assumption that technology flows not only bring outdated blueprints, but are also a source of new technological development. Thus, catch-up is viewed as a process in which poorer countries both imitate and adapt older technology.

Theories of technology gaps incorporate Posner’s and Vernon’s theories on economic development (Posner, 1961; Vernon, 1966) into a Schumpeterian view on innovation and imitation. The idea is that new technology is developed in certain countries that are constantly at the technological forefront. Later on in the product cycle, production is relocated to other countries. This may be the effect of two independent factors. First, as the advanced country keeps on innovating, efficiency and wages increase. Therefore productions of some goods become unprofitable. The relocation of the Western European textile industry to low-wage countries is an example of this mechanism. Second, as a technology grows old, it becomes well known. The technology changes nature from a semi-private to a public good. As a consequence of these conditions, other countries further down on the productivity ladder take over production of the older goods. Krugman (1979) formalises the diffusion effect; in a later study (Krugman, 1986) he analyses the crowding out effect. Barro and Sala-I-Martin (1997) present a model in which technology gaps and diffusion of technology to poor countries are combined with endogenous innovation in the leading country. An interesting implication in these models is that increased productivity at the forefront is always of benefit to both rich and poor countries. Diffusion of knowledge, however, benefits poorer countries but not necessarily richer ones. Global intellectual property rights, imposed on all countries through the WTO’s Trade-Related Intellectual Property Rights agreement (TRIPS), are an example of how global politics have traded off these considerations.6

---

6 The TRIPS agreement obliges all member countries in the WTO to establish a standardised patent institution. In principle, all patentable innovations can be patented in all countries. Maskus (2000) discusses the consequences of the TRIPS agreement. Helpman (1993) extends the technology gap model of Krugman (1986) with a discussion of intellectual property rights and their implications for the ability of poor countries to catch up.
In Fagerberg (1988) growth in a set of countries is assumed to be a function of technological distance between the country in question and the world economic leader (the US) and of the resources devoted to increasing the country’s absorptive capacity. Fagerberg demonstrates that the outcome of economic development might be both convergence and divergence. He proposes that a country’s income level will depend on its own R&D, diffusion of knowledge from abroad and the country’s capacity to exploit foreign knowledge. The technology gap hypothesis is that countries lagging far behind the frontier have a larger potential for catch-up than other countries. The frontier is taken to be knowledge in the leading economy in the world. In this set-up, therefore, growth will depend on a country’s initial income (which indicates the technology gap), its absorptive capacity and possibly some other variables. The empirical implications of this model are very similar to the empirical formulations of the neo-classical growth model. In the technology gap models, poor countries are predicted to have a high potential for growth through technology imports; in neo-classical models, they are predicted to grow fast because of high returns to capital.

Verspagen (1991) models catch-up and technology flows in a similar way. He explicitly allows for the existence of underdevelopment traps. In the case of countries that are way behind the technological leader, their ability to make use of technology flows is limited. Other countries, further up the productivity ladder, have higher absorptive capacity and are able to keep the technology gap constant or reduce it. Thus, Verspagen’s model predicts a world in which there is a club of very poor countries and another club of converging wealthy countries.

**Summing up**

Recent growth theory is to a less extent than traditional theory based on assumptions of decreasing returns on physical or human capital. Leaving behind that assumption also implies that the traditional source of convergence vanishes. In a large class of models, convergence in income per capita is shown to be dependent on whether technology flows are global or local in scope and whether knowledge spills over between industries. Moreover, when there is international trade, convergence depends on the extent to which prices of goods imported from technological leaders tend to fall over time as technology progresses.
3. Empirical Evidence

Measurement and methodology
In the empirical literature several measures of convergence have been proposed. The first has already been mentioned: the lack of an unconditional systematic relationship between the initial level of GDP and subsequent growth rates for the world economy is referred to as unconditional $\beta$-divergence. Conditional $\beta$-convergence is the occurrence of convergence when other factors are controlled for. $\beta$-convergence, therefore, denotes a negative coefficient for initial level of GDP in a cross-section regression on growth rates for a sample of countries according to the regression equation:

$$g = \frac{1}{T} \log \left( \frac{y_{it}}{y_{i,t-T}} \right) = \alpha + \beta \log(y_{i,t-T}) + \gamma X_{i,t-T} + u$$

Above, $y_{it}$ denotes GDP per capita in entity $i$ at time $t$. $T$ denotes the time from the initial year to the last year; $u$ is the regression residual. The regression equation above therefore expresses the hypothesis that growth depends in the (log of) initial income and a set of other variables. One distinguishes between conditional $\beta$-convergence and unconditional $\beta$-convergence according to whether other relevant variables, denoted by the vector $X$, are included or not. Unconditional $\beta$-convergence means that $\beta$ is negative and significant when $X$ is left out. Conditional $\beta$-convergence means that $\beta$ is negative and significant when other explanatory variables are also included in the regression. The literature is not conclusive on what variables to include. Often included are variables reflecting openness to trade, the population’s educational level and the level of investments. Levine and Renelt (1992) and Barro (1997) provide critical reviews of what conditioning variables to include in cross-country growth regressions. Dobson et al. (2001) argue that the rate of convergence obtained in such regressions depends on what conditioning variables are included. In the following subsection, I give an overview of some empirical results in this tradition of cross-sectional studies of economic growth.

The reader should note that the above expression might capture the neo-classical hypothesis of convergence, the endogenous growth hypothesis with international technology diffusion and the technology gap models (when a lag to a frontier is included).

A more restrictive version of convergence is so-called $\sigma$-con-
σ-convergence denotes that the standard deviation of (the log of) GDP per capita in a sample of countries decreases over time. σ-convergence is a stronger criterion than β-convergence in the sense that absence of σ-divergence can co-exist with β-convergence but not the other way around. The relation between β-convergence and σ-convergence may be derived from the above equation. Rewriting it and setting T=1, a equation of log(yₜ) is obtained. This equation is a so-called difference equation in which the level in one period depends on the level in the previous period. The term u remains in the equation and is assumed to be a random variable with zero mean and constant variance over time and over our units of observation (absence of autocorrelation and heteroscedasticity). Taking the sample variance of this expression gives:

\[ \sigma_{yt}^2 = (1 + \beta)^2 \sigma_{yt-1}^2 + \sigma_u^2 \]

Above, \( \sigma_{yt}^2 \) denotes sample variance of the log of GDP per capita in year t and \( \sigma_u^2 \) is the sample variance of u. It is seen that the expression for variance in GDP levels per capita is a function of β. If β is negative (as implied by the β-convergence hypothesis), it contributes to reduced sample variance over time. Variance might nevertheless increase if the contribution from the error term, u, is larger than the contribution from β-convergence.

The second tradition of empirical studies I will review is the analyses of total factor productivity. From the production function presented earlier we have:

\[ Y = AK^{\alpha}L^{1-\alpha} \]

\[ \frac{\dot{A}}{A} = \frac{\dot{Y}}{Y} - \alpha \frac{\dot{K}}{K} - (1 - \alpha) \frac{\dot{L}}{L} = \gamma X + u \]

From the second of these equations, growth in total factor productivity is expressed as the difference between growth in total levels of GDP and a weighted average of factors of production (capital and labour in this simple stylised example). This is similar to the expressions above. The last of these expresses the hypothesis that growth in total factor productivity is a linear function of possible explanatory variables.

Normally it is assumed that factors of production are paid their marginal productivity. This means that capital and workers are em-
ployed until the value of the extra production that results equals the cost of hiring them. In this case workers’ share of production is equal to \((1-\alpha)\) and capital’s share of production is equal to \(\alpha\). Therefore, growth in productivity will be equal to the difference between the growth rates in GDP and the reward to the factors of production, times the growth rates in these. In many countries both investments and wages are observable. Given the above assumptions, total factor productivity can be estimated. Growth in factor productivity is used as a dependent variable in this kind of study. This is the growth-accounting procedure.

Studios of total factor productivity have revealed that growth in total factor productivity is substantial. In fact, several studies have demonstrated that productivity growth accounts for the major share of growth. Growth in total factor productivity has been denoted a measure of our ignorance (Abramowitz, 1956) because it is the share of growth that cannot be accounted for by growth in traditional factors of production. In recent research, however, it is very often the productivity that is subject to research.

There are important limitations to growth accounting and studies of total factor productivity. This approach is based on assumptions of constant returns to scale in production and of perfect competition. Barro and Sala-I-Martin (1995) also point out that growth in capital and production might be the consequence of growth in total factor productivity. If so, the usual measures of total factor productivity underestimate the contribution from technological change and overestimate the contribution from capital accumulation.

**Empirical results – an overview**

*Growth regressions*. Growth regressions have been very popular in recent years. There are two traditions of growth regressions on data sets for global data. The first attempts to test the neo-classical growth model, often extended with human capital. These studies indicate that the level of GDP per capita can be well explained only by inclusion of investments and human capital (see, for example, Mankiw *et al.*, 1992). However, these two variables do not succeed in explaining growth, i.e., changes in levels, very well.

The second approach to global data sets has been to include a large set of explanatory variables in regressions on growth. These exercises have been useful in at least two senses. First, they reveal possible explanations for growth. In growth regressions, investments, schooling (male, but not female!) and initial income are robust var-
variables correlating with growth in many studies. Resource abundance is negatively related to growth, at least in countries with institutions of low quality (Sachs and Warner, 1995b and Mehlum et al., 2002). Political instability is detected as important for economic growth, although it is hard to determine whether this reflects the impact of social unrest, insecure property rights or lack of other institutional qualities. Openness to trade is a less robust explanatory variable, but several studies indicate a strong, but not very significant effect. Rodriguez and Rodrik (1999) present evidence of the opposite: they find no significant relationship between trade policy and economic growth. Second, growth regressions clarify the concept of convergence: by use of such regressions on different samples of countries and with different explanatory variables, one may detect to what extent initial income robustly influences subsequent growth.

Growth regressions are not without problems. I will emphasise three of them. First, it is not clear what the direction of causation between the explanatory variables and growth is. Neither is it clear what variables to include in growth regressions. Levine and Renelt (1992) have constructed a test for the robustness of explanatory variables in growth regression. The essence of their test is that a variable should be statistically significant and of the same sign in regressions independently of the inclusion of other different variables. This implies that investment will be a robust explanatory variable if regressions give a positive and significant result independently of whether other variables, like schooling, trade policy etc. are included. Second, growth regressions of the type cited below very often presume that countries can be observed independently. The most common regression methods are based on ordinary least squares regression and it is not taken into account how countries interact with each other. Third, growth regressions have limited explanatory power. One reason for this is that regressions on the largest samples possible provide researchers with a small set of available explanatory variables. We believe that investments in R&D are an important source of growth, but for many countries R&D data are not available. Investments in human capital are therefore often approximated, for instance, by data on school enrolment.

---

7 The finding that female schooling is associated with low growth rates has spurred debate and further studies. There is now agreement that it should not to be interpreted as a causal effect, but it represents a puzzle in the data. See Klasen (2002).

8 It is found in several studies that a high share of exports of natural resources in a country's GDP is negatively correlated with economic growth. Two main explanations have been put forward. The first is Dutch decease, which implies that other productive activities are crowded out. The second is that resource abundance stimulates rent-seeking activities.
Table 1 reports results from some important studies. It is seen that only three variables stand out as robust explanatory variables of growth. These are initial income, investments and international trade. Other variables are often not significant or their significance (and even their sign) depends on what other variables are included. Often variables seem to have non-linear effects. This is the case both for indexes of democracy and for inequality.

Table 1. Estimation results for growth in GDP per capita, global data, results from various studies

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reference</th>
<th>Effect</th>
<th>Robust/Fragile</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) initial income</td>
<td>Ba, MK, B, BS, I, SW(1), M</td>
<td>-*</td>
<td>R</td>
</tr>
<tr>
<td>(2) investments</td>
<td>Ba, MK, B, BS, I, S</td>
<td>++</td>
<td>R</td>
</tr>
<tr>
<td>(3) human capital</td>
<td>Ba, MK, BS, I</td>
<td>+ (-)*</td>
<td></td>
</tr>
<tr>
<td>(4) trade</td>
<td>FR</td>
<td>+*</td>
<td>R</td>
</tr>
<tr>
<td>(5) trade policy</td>
<td>S, SW (1), RR</td>
<td>+(.)*</td>
<td>F</td>
</tr>
<tr>
<td>(6) foreign direct investments</td>
<td>BLZ</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>(7) corruption</td>
<td>MK</td>
<td>-*</td>
<td></td>
</tr>
<tr>
<td>(8) democracy</td>
<td>B</td>
<td>±*</td>
<td></td>
</tr>
<tr>
<td>(9) health</td>
<td>B</td>
<td>+*</td>
<td></td>
</tr>
<tr>
<td>(10) inequality</td>
<td>B(1), PT</td>
<td>±*</td>
<td></td>
</tr>
<tr>
<td>(11) inflation</td>
<td>LR</td>
<td>-</td>
<td>F</td>
</tr>
<tr>
<td>(12) regions</td>
<td>SM, B,</td>
<td>+* (East Asia, lat)</td>
<td></td>
</tr>
<tr>
<td>(13) rule of law</td>
<td>SM, B,</td>
<td>+*</td>
<td></td>
</tr>
<tr>
<td>(14) religion</td>
<td>SM</td>
<td>-(±)*</td>
<td>(Christianity)</td>
</tr>
<tr>
<td>(15) size of public sector</td>
<td>Ba, BS</td>
<td>-*</td>
<td></td>
</tr>
<tr>
<td>(16) resources</td>
<td>SW, M</td>
<td>-(±)*</td>
<td></td>
</tr>
<tr>
<td>(17) political stability</td>
<td></td>
<td>+*</td>
<td></td>
</tr>
</tbody>
</table>


Table 1 reports results from some important studies. It is seen that only three variables stand out as robust explanatory variables of growth. These are initial income, investments and international trade. Other variables are often not significant or their significance (and even their sign) depends on what other variables are included. Often variables seem to have non-linear effects. This is the case both for indexes of democracy and for inequality.

*Studies of total factor productivity.* In empirical studies of factor productivity convergence is not the issue. The focus is on productivity and its determinants. In studies like these, a hypothesis that is often tested is the predicted potential for lagging countries,
sectors or firms to catch up in terms of productivity by use of knowledge developed elsewhere. In order to study the effects of innovation and knowledge flows or spillovers, as modelled in endogenous growth models, many researchers have chosen to focus on smaller data sets for which more variables are available. Such variables are data on R&D, patents and, most important for our subjects, those reflecting the diffusion of technology. I will distinguish between findings of embodied and disembodied technology flows since their interpretations differ.

As discussed above, technology flows potentially have many forms. One is technology flows embodied in goods. Buyers benefit from the knowledge that is used to develop a good, both if the good is used as a factor in production and if it is used for consumption. A set of studies has revealed important effects of embodied technology flows for growth in factor productivity.

(A) Coe and Helpman (1995) hypothesise that growth in productivity depends positively on a country’s own R&D and other countries’ R&D. They assume that others’ R&D is imported through imports of capital goods. They therefore regress productivity growth in the OECD countries on each country’s own R&D and a weighted sum of other countries’ R&D where the weights are the shares of imports from those countries to the country in question. The results are striking: Coe and Helpman find that most productivity growth results from foreign R&D and not from national R&D. The import of foreign R&D has greater influence on smaller countries than on large ones. A later study is that of Frantzen (2001), who extended Coe and Helpman’s analysis to a longer period.

(B) Coe, Helpman and Hoffmeister (1997) extend the above study to a group of developing countries. In this study, evidence is found that foreign R&D stocks and imports of capital goods from other countries explain growth in total factor productivity more than does, for instance, schooling. Furthermore, the effect of foreign R&D seems to be larger the more open the economy is.

(C) Lichtenberg and van Pottelsberghe de la Potterie (1996) aim at extending the analysis by Coe and Helpman to flows of international foreign direct investments. Their findings do not lend support to the view that important technology flows from the investing country to the recipient country. Their findings suggest the opposite; the investing country benefits from R&D in the
host country. This is an important finding since many proponents of foreign direct investments (FDI) argue that inward FDI are an important source of technology. In fact, several studies indicate that this is not the case.

(D) Another extension of Coe and Helpman is the study by Eaton and Kortum (2001). They hypothesise that imports of capital goods depend on these goods' prices. They estimate a price index of trade capital goods for importing countries and find that countries that face high prices of imported capital goods experience lower productivity and income per capita.

(E) Similar results are found in Maurseth (2003) in which a theoretical price index of capital goods is constructed. The price index is constructed according to an assumption that geographical distance is an important barrier to trade. Therefore, the price index of capital will be higher in peripheral countries. This gives an explanation for the empirical regularity that high-income countries are located near large markets. Similarly these results indicate that growth rates between countries will be geographically clustered. Negative growth in one country infects neighbouring countries and positive growth is likewise transmitted to neighbours.

(F) These results are in line with Easterly and Levine (1998) who explicitly estimate contagion effects in economic growth. They find that countries are affected by the growth destiny of their neighbours. For the world economy, the nature of these findings contribute to explanations of why clusters of countries get rich and other clusters remain poor.

There is also another set of studies that focuses on disembodied technology flows. These denote flows of technology that occurs without economic transactions as prerequisites. Examples of such technology flows are exchange of knowledge in academic research, industrial espionage and reverse engineering.

(A) In models of technology gaps, the main hypothesis is that a technology gap between a poor country and the leading country potentially favours growth in the poor country. Fagerberg (1987) demonstrates that, for a sample of 25 countries, including the OECD countries, growth is well explained as a positive function
of each country’s number of patents (as a measure of innovation), a negative function of the technology level (measured as the country’s own GDP per capita) and investments. The negative coefficient of initial level of GDP is interpreted as a technology gap between the country in question and the technology leader in the sample (the US). It should be noted that this study does not differ from growth regressions except for the inclusion of patents as an indicator of technology. The interpretations of the result differ, however.

(B) In the same vein, Griffith, Redding and Van Reenen (2000) estimate productivity in industries in a country as a function of the lag between productivity of the industry in this country and the productivity of the same industry in the country with the highest productivity in that industry. They find clear evidence of convergence in productivity levels between countries.

(C) Eaton and Kortum (1996) analyse international patenting. They hypothesise that if an invention is patented in a country (particularly when it is not where the invention originated), it signals a transfer of technology. They estimate the determinants of international patenting and find, among other things, that distance reduces knowledge diffusion. They find positive and significant effects of international knowledge flows in the same vein as Coe and Helpman (1995): foreign innovation is more important than national innovation in smaller countries. Eaton and Kortum analyse growth in labour productivity, however.

(D) Keller (2002) estimates total factor productivity as a function of a country’s own R&D and that of others, in 14 countries, but for a large set of industries. He finds that the effects of others’ R&D on a sector’s productivity decrease rapidly with geographical distance and linguistic borders.

(E) Verspagen (1997) estimates total factor productivity in different industries and uses patent citations as the weights for technology diffusion from one sector to another. Verspagen’s analysis, there also seem to be important effects of technology diffusion. The same result is found in Maurseth (2001) for a disaggregated set of Western European regions.

To sum up: studies of total factor productivity suggest that the pro-
ductivity in industries and countries depends to a large degree on technology flows from other sources rather than from their own inventions.

**σ-convergence and other types of distribution dynamics.** As mentioned above, a strict test of convergence is σ-convergence. σ-convergence denotes reduced standard deviation in the cross-country income distribution over time. As such the measure is extremely simple. There have been only a few studies that incorporate explanatory variables in analyses of σ-convergence. Two of these are Ben-David (1996) and Ben-David and Kimhy (2001). Ben-David acknowledges the problems of including trade in growth regressions. He therefore analyses σ-convergence among trading partners. In particular, he finds that pairs of countries that trade intensively with each other show less dispersion in their income than other countries. Similarly, he finds that pairs of countries that increase their trade relations, experience reduced dispersion in their income per capita. A related finding is presented in Figure 3. The figure shows the dispersion in income per capita among countries standardised to world average and in income per capita standardised to a distance-weighted world average. In analyses of geography in general (and for economic growth in particular), the hypothesis is that some variable x in entity i influences some variable y in entity j as a decreasing function of the distance from i to j, $d_{ij}$. Therefore, a distance weights matrix was constructed according to:

\[
W_{ij} = \frac{1/d_{ij}}{\sum_{j=1}^{n} 1/d_{ij}}
\]

The resulting weight matrix postulates that the influence of any variable between two countries decreases with the inverse of the distance between them. The weights are standardised so that they add up to one for each country. This makes it easier to construct weighted averages of variables for countries.

The figure reveals that dispersion is less between neighbours but that $s$-divergence occurs in both the overall distribution and the distance-normalised distribution.
Quah (1993 and 1996) argues that both $\beta$-convergence and $\sigma$-convergence are crude measures of convergence. For instance, both can be consistent with Baumol’s notion of convergence clubs, in which there are clubs of countries converging towards common levels of GDP per capita (Baumol, 1986). Quah proposes to report transition probabilities from percentiles of the distribution of income over time. Thus, growth clubs would be characterised by more entries into certain percentiles of the population than exits from the same percentiles. The essence of his proposal is demonstrated in Figure 4, which graphs the ranking of 104 countries in the world economy in 1960 and 1990. Quah’s transition probabilities correspond to countries jumping from one of the graphed squares to another. He characterises the cross-country income distribution as stable if countries remain within those squares and unstable if they jump out of their squares. Figure 4 demonstrates that the

---

9 An objection to this approach is that countries at the lower and higher parts of the income distribution can only ‘jump’ in one direction. Still, comparing the ranking of countries at different points in time gives an impression of stability versus instability.
income distribution across countries in the world was more stable for rich countries than for others. This reflects the clear convergence among the rich OECD countries.

4. Summary and Conclusion

Whether countries will tend to converge in income per capita is an important question for students of economic growth. While convergence was an inherent prediction in the traditional neo-classical growth model because of decreasing returns to capital, in recent theories convergence is predicted to depend on diffusion of knowledge. Diffusion of knowledge takes many forms and is often distinguished as being embodied in traded goods and disembodied flows of knowledge.

Recent empirical research lends support to the neo-classical hypothesis of conditional convergence: when other relevant factors are accounted for, there is convergence in GDP per capita. It is not clear from growth regressions what to conclude from this. One
interpretation is that this supports the hypothesis of decreasing returns to capital. Another is that low levels of initial income indicate a large potential for catch-up through assimilation of technology. It is important that conditional convergence is not equivalent with a collapsing or narrowing income distribution. In fact, differences between rich and poor countries have increased. Growth regressions have revealed important potential sources of growth, however. These are investments in human and physical capital, institutional quality and openness to trade.

Studies of smaller datasets demonstrate a potentially large influence of technology diffusion. Of the channels for knowledge spillovers, trade between countries has been identified as important. It is not clear from recent studies whether trade-induced spillovers dominate in importance over the disembodied spillovers analysed in the first generations of endogenous growth models.

While growth economics has revealed a set of important mechanisms related to economic growth, it has not resulted in a toolbox for growth-promoting policies. In particular for very poor countries, there are many remaining questions. The effects of institutional quality, governance and geography on economic growth seem to be major issues for future research.

References


Penn World Tables, Mark 5.6 is available at http://pwt.econ.upenn.edu/
Developing Countries and the Tourist Industry in the Internet Age: The Case of Namibia

Arne Wiig

1. Introduction

For low- and middle-income countries, tourism represents almost 8 per cent of merchandise trade. Around one-third of international tourism is captured by developing countries, and tourism is the only major sector in international service trade where developing countries have consistently had surpluses. Worldwide, travel services constitute 32 per cent of service exports, but in Africa they constitute 56 per cent. It could be claimed that these high shares of service exports reflect the low level of the other service sectors rather than a high level of tourism. This is only partly correct. International tourism has shown significant growth rates and these were 50 per cent higher in developing countries than developed countries in the period 1980–96. From 1995 to 1999 annual tourism in African countries increased by 6.9 per cent compared to a world average of 2.9 per cent (WTO, 2000a). In the past, high growth rates have led to an upsurge in the importance of tourism, a proc-

Note of acknowledgements. A previous version of this article was presented at the annual conference of the Norwegian Association of Development Research (NFU) in Trondheim, 14–15 Nov. 2002. I am grateful for comments from Henry DeGroot, Hildegunn Nordaas, Ivar Kolstad, Gaute Torsvik and Ingvild Hestad. Comments from referees are also acknowledged. Any remaining errors are mine alone.

I received financial support from the Norwegian Research Council and NORAD. Fieldwork in Namibia was financed by NORAD.

1 38 per cent of tourist arrivals and 31 per cent of tourism receipts are accruing to low- and middle-income countries of which China, Thailand, Mexico, Malaysia, Singapore and Indonesia are the most important tourist destinations (World Bank, 2003:354).
ess that seems to be continuing. UNCTAD (2001ab) proclaims that there is further tourism potential for developing countries. In fact, tourism development appears to be one of the most valuable avenues for reducing the marginalisation of least developed countries (LDCs) in the global economy.2

The tourist industry represents an interesting avenue for poverty reduction and economic development (ODI, 1999; Sinclair, 1998). Tourism is labour-intensive; there are potential links with other industries such as construction, agriculture and transport; and the poor may have access to tourist assets such as natural resources and culture. At the same time, tourism may lead to de-industrialisation (undermine other industries) if prices of non-tradeables increase significantly (Copeland, 1991).3

My point of departure is that tourism is an information-intensive bundle or package of goods. As such, the tourism sector is intertwined with the development of information and communication technology (ICT), a factor that is hardly mentioned in the literature referred to above.

ICT already plays a significant role in tourism. Tourism is one of the most important sectors that apply ICT and one would expect that changes in ICT, particularly the arrival of the Internet, to influence the structure of the industry. UNCTAD (2000) proclaims that e-commerce and ICT represent an opportunity for developing countries to improve their relative position through a significant reduction in transaction costs. E-commerce permits development through productivity gains in supply chain management and is expected to represent about 20 per cent of worldwide business-to-business and retail transactions by 2006 (UNCTAD, 2002a). The Internet and e-commerce represent promising avenues for further development of the tourist industry in developing countries and challenge the traditional role played by ‘bricks and mortar’ intermediaries. Yet, one cannot be too optimistic. The ‘new economy’ is not so new, in the sense that fundamental economic principles also apply to it as much as to the ‘old economy’. We do not yet know the fate of the traditional intermediaries and whether there is room for new agents, for example from developing countries. Traditional agents may as well

---

2 The participation of LDCs is small and uneven. Tanzania, Maldives, Cambodia, Nepal and Uganda account for half of all tourist receipts by LDCs.

3 As argued elsewhere, in the case of Namibia, this is not really a problem since manufacturing already plays a minor role, partly due to high labour costs (Brits and Wiig, 1998).
exploit the new opportunities created by the ICT revolution and adapt existing business activities to the new circumstances.

This article analyses how the use of ICT in the tourist industry influences the structure of the supply chain and looks at the potential impact on market share and revenue. For most developing countries, the retention of tourism revenue is currently as low as 30 per cent (WTO, 1998a). Most analyses focus on import leakage (part of tourist consumption is imported goods) and factor income leakage (airline companies, hotels, travel lodges and car rental firms are in many cases owned by foreigners). My focus is on savings in transactions costs. I will distinguish between three different consequences of ICT. The Internet may lead to a reduction in distribution costs. The first question I raise is whether the Internet makes it possible for a service provider to reach the customer directly through marketing. Normally, consumers buy their tourist goods or services through a travel agent in their home country. The agent receives a significant commission from the supplier, but the commission varies according to the tourist product. Through direct marketing the service provider saves distribution costs, which in turn may increase the market share and the retention of revenue. The second mechanism analysed is the Internet’s impact on competition. The question I raise is whether the Internet increases competition between intermediaries. My focus is on competition across networks of global distribution systems (GDSs). Increased competition may reduce the fees the network charges and make it more likely for an individual service provider to connect to a GDS and thereby reach a larger market. The third mechanism is the Internet’s capacity to market various marketing destination messages. The main question addressed is how co-operation between service providers can reduce distribution costs and exploit network externalities by creating destination portals.

While we know that online travel constitutes about 10 per cent of the market in developed countries, we hardly know anything about the extent of ICT and e-commerce or the importance of intermediaries in developing countries.\textsuperscript{4} Statistical data are important in order to design strategies for tourism development, and I will address the three questions raised above in relation to ICT with data from the Namibian tourist industry.

\textsuperscript{4} UNCTAD (2002b:7) provides an overview of available statistics, but no references are given to data from developing countries.
This article is organised as follows. As a background to the analysis, Section 2 elaborates on the information-intensive character of the tourist industry and discusses spillovers. Section 3 analyses the three mechanisms in which ICT may influence the supply chain in the tourist industry described above. Section 3.1 analyses conditions under which the service provider may substitute for the travel agent. Particular emphasis is given to the travel agent’s role as a bundler and a ‘certifier’ of goods and services. The next two subsections analyse the impact of ICT on competition between distributors and on co-operation among service providers respectively. Based on the critical variable defined in Section 3, Section 4 presents results from a survey among stakeholders in the tourist industry in Namibia. Section 5 concludes.

2. Tourism – an Information-Intensive Product

The following section elaborates on the characteristics of the tourist product as an information-intensive bundle of goods and services. I then look at the wider perspective, in particular spillovers into other sectors such as telecommunications, financial markets and factor markets (i.e. labour, markets and skill formations), which underlines the importance of complementary goods and services. In order to analyse the impact of ICT on the supply chain in the tourist industry, corresponding changes in complementary assets must be analysed.

2.1 Characteristics of the tourist product

Tourism is a composite bundle of goods and services geographically segmented across specific geographical areas. It includes accommodation, transport facilities (air transport and vehicles), activities (what tourists do during their stay), attractions and ancillary services (banking, telecommunications, hospitals). Factors such as the destination’s security level, confidence and trust between the main actors, and cultural identification are also important factors influencing the value of the tourist product. Of particular relevance in a Namibian context are wildlife and spectacular landscapes (Ashley, 1998; Brits and Wiig, 1998).

Consumers often do not place a separate value on individual elements of the composite goods. There are complementarities between certain elements. Site-specific assets such as spectacular scenery or wilderness may, for example, increase the ‘value’ of the
accommodation. Similarly, ancillary services and trust make it more attractive to be a tourist in a particular place. The consumer’s choice of a particular destination and tourist product is determined less by individual components such as hotel facilities than by an attractive combination of various elements.

Tourism is an information-intensive industry, particularly during the booking phase (UNCTAD, 2001b) and when new tourist destinations are introduced. Producers and intermediaries try to earn the confidence of their customers through providing quality information. On delivery, tourism is labour-intensive. But also at this stage, provision of information about local history and culture influences the value of the product. Through tourism, people are exposed to different cultures, and informed tourism may change stereotypical images (which to some extent are also information goods), which have impeded travel activities, trade and direct foreign investments.

To reach international tourists, suppliers need intermediaries such as travel agents and tour operators who can obtain and provide information and bundle goods. Travel agents have traditionally taken care of these tasks through the use of different computer reservation systems (CRSs). During the 1980s, airline companies’ computer reservation systems became global distribution systems (GDSs).

The travel agents and tour operators are also important certifiers of goods, which is of particular importance for new tourist destinations where consumers have no prior experience. In addition, there is a legal aspect to be considered. Consumers may sue travel agents if they sell substandard products. It is more difficult to sue individual service suppliers in remote areas that lack a sound legal system. Inclusive tour charter is therefore a common way to travel to new tourist destinations.

The main mode of supply is movement of consumers – not movement of producers or products. As with other service industries, foreign direct investments also play a significant role.

2.2 Spillovers
Tourism generates much employment worldwide. One person in ten is employed by the tourist sector, and tourism is a significant sector for small- and medium-scale enterprises (SMEs). Countries in the World Tourist Organization are in the process of creating satellite accounts for the tourist sector. Such accounts make it possible to analyse the importance of tourism in a particular country
and its indirect impact on other sectors (indirect multipliers). Satellite accounts are not generally available for developing countries.

In Namibia, tourism plays an important role in terms of both employment and exports, particularly for SMEs located in remote areas of the country where diversification is needed. More than 20,000 people are employed directly or indirectly in the tourist industry (see Government of Namibia, 1997). The indirect multiplier is estimated to be 0.7 although no input-output matrix is available. From visitors’ surveys, it is estimated that air fare and accommodation constitute around 70 per cent of the total package costs and that the local carrier Air Namibia serves 50 per cent of travellers to Namibia. A similar pattern is found in developed countries.5

In addition to generating employment, tourism may also enhance a country’s social capital through skill-formation. A positive service attitude, fluency in the language spoken by the tourists and a rating system of service providers may increase the quality of the product. Certain labour and quality standards are necessary in order to compete internationally. Such standards upgrade the quality of the staff and services provided, not only in tourism but also in other sectors.

The composite character of tourism may also lead to more cooperation. Ashley (1998) found that tourism in communal areas stimulated co-operation between community members as well as co-operation across communities.

Many developing countries market tourism as scenery- or wildlife-intensive, making it of utmost importance to preserve the wildlife. With a proper regulation regime, income from tourism may facilitate conservation and investment in wildlife and habitat, and also sustain culture (Ashley, 1998 and 2000). Namibia, for instance, is marketing its product as a ‘wildlife product’ and without a proper conservation policy tourism would not be sustainable. The focus of this article, however, is on tourism as an activity that may enhance information and communication technology (ICT) and financial services. These services have a spillover impact on local communities as well as at the national level and may lead to economic growth. In OECD countries, Roller and Waverman (2001) found

---

5 For instance, in Norway passenger transport services (bought in Norway) constitute nearly one-third of the total tourist expenditure. However, in Norway, food and beverages – not accommodation – constitute the second most important tourist expenditure, at 20 per cent of the total.

(See http://www.ssb.no/english/subjects/09/01/turismesat_en/tab-2002-08-29-02-en.html)
that ICT development stimulated economic growth through an increase in the total factor productivity.\(^6\) This result has been disputed in a developing country context (see, for instance, Aochamub et al., 2002), but find some support in Jacobsen (2003).

A well-developed telecommunication sector is a prerequisite for international communication and supply chain management, including the development of e-commerce. Without Internet access, it is impossible to create web pages and have effective dialogue with partners or customers. Customers will be faced with high search costs when they look for tourist services in remote areas. With Internet access, they can be only a click away. In addition, financial services (online banking facilitates or similar services) are needed to encourage e-commerce.

The tourist sector plays an important role in increasing the demand for ICT services. Tourism is generally an ICT-intensive industry in terms of registered home pages, numbers of visitors to websites and online trade. In fact, tourism (travel, transport and hotel reservation) is the single most important product bought online, and constitutes 38 per cent of all online trade (UNCTAD, 2002a). Tourism represents the product with the highest growth rate (40 per cent) in e-transactions.\(^7\) According to the consultancy firm Forrester, more than 60 million households in the United States booked travel online during 2002, spending approximately US$ 20 billion, constituting around 10 per cent of the travel market. As measured by the total number of composite visitors, online travel agents and air companies were the third most important websites visited (after Yahoo, and AOL) in 2001.\(^8\) There are also more visitors to information resources such as Lonely Planet and Milesource than to those on hotels and rental cars.

3. The Internet and Industry Structure
Lower transaction and distribution costs through the Internet may change the way the supply chain is managed, particularly the role played by intermediaries. In Section 3.1, I will focus on how the Internet makes it possible to provide information services directly to the customers and thereby reduce the role of travel agents. In

\(^6\) The increase in growth that cannot be explained by added labour and capital.
\(^8\) http://www.top9.com/top99s/top99_web_sites.html. The last available figures are from March 2001.
that section it is assumed that the number of GDSs (or networks) and the number of firms connected to a network is fixed. Section 3.2 extends the analysis to discussions on how the Internet may influence competition between networks and the individual service providers’ incentives to participate in these networks. Section 3.3 focuses on how service providers may co-operate by creating a particular type of network, the so-called destination marketing portal.

3.1 The Internet reduces the importance of the travel agent
The traditional structure of the supply chain in the tourist industry consists of consumers, travel agents, CRS suppliers and service providers. In addition, local and overseas tour operators operate as merchants. Consumers undertake private or business travel, which are two distinct market segments. Our focus is on the private tourist market. Customers differ according to their willingness to pay for the tourist bundle, their attitudes to risk and their computer skills. These are all factors assumed to influence the demand of tourist services and the choice of distribution channel. Consumers may either buy the tourist bundle directly through individual service providers or through intermediaries. Most interactions in the travel market are currently completed through intermediaries.

Travel agents and tour operators are both intermediaries who through their services facilitate the exchange of tourist services. Traditionally, the travel agent is seen as a retailer and the tour operator as a wholesaler. But this is only partly correct. To some extent, the tour operator and the traditional retailer play similar roles; both assume ownership of the product. Operators are remunerated through residual surplus (the difference between buying and selling prices). The travel agent, on the other hand, is a broker, does not own the product and is not faced with the risk of perishable products. The travel agent is remunerated by commission. Both agents provide information services about the supply and demand structure for a bundle of individual products which make up the composite good of tourism. In addition to creating a package, these intermediaries also serve as certifiers (guaranteeing quality and payment). In the following, I will restrict the analysis to travel agents since they, as brokers, are expected to be challenged the most by the new technology.

9 See Spulber (1996) for an overview of the theory of intermediaries.
10 Intermediaries also provide insurance and liquidity, and facilitate logistic management (Brousseau, 2002; Spulber, 1999).
Consumers do not have direct access to a network of information services, while travel agents do through their connections to CRSs or GDSs. Travel agents can use many GDSs, but normally use only one or two since each system requires a specific technology and travel agents normally get a discount after a certain volume of reservations. The agents make requests regarding connections, availability of seats, hotels, etc., through their network of information services. For an individual service supplier to be selected, the supplier needs either to have a direct contact with the travel agent, some system of electronic data interchange (EDI) or to be connected to a GDS. When connected to a GDS, their inventory is available to all subscribing agencies worldwide within the GDS they participate in.

Figure 1 reflects how orders are placed and money transactions made in the traditional industry structure. The travel agent is the customer of a GDS and pays for its services, and gets commissions from the individual service suppliers. Individual service providers normally pay a fee per transaction (segment) to a GDS, but also need to make fixed investments in computer systems.
The Internet has introduced new distribution channels. Services may be offered directly to the customers either by the individual service provider on a home page or by ‘virtual travel agencies’ providing information through the network. Consumers may access information by searching the Internet and respond interactively, if full-fledged e-commerce solutions are available. The travel agent is therefore challenged both by direct contact between the consumer and the service supplier, and by virtual travel agencies. The increased pressure from suppliers is illustrated by air companies’ recent moves to reduce the commissions paid to travel agents.

In Figure 2, below, I have sketched the supply chain (without the travel agent) including electronic intermediaries. In the figure, a distinction is made between four types of virtual agency. These are portals connected to existing GDSs, other virtual travel agents, travel portals of a group of individual service providers (air companies) and national portals created at particular destinations. In this section, I will focus on the first category only, emphasising a situation where competition between existing electronic intermediaries is unchanged, but services can be offered through the Internet. The alternative online option is direct marketing (which may be termed the fragmented solution). In the following section, I will discuss competition across networks.

The Internet has made it possible to make global distribution networks accessible for customers on the Internet, as Sabre has done by establishing Travelocity. Whether electronic intermediaries will take over the role played by traditional travel agents depends on whether electronic intermediaries are able to provide the same services at competitive prices, in terms of providing information, creating attractive packages and winning the trust of customers. The cost of providing information is low, since the only difference is that information previously provided by GDS systems is now made available on the Internet by the same agents. A likely implication is thus that virtual travel agencies will be able to provide information efficiently. Information goods have low marginal costs, the cost of creating a package of services is low, and virtual agencies are able to exploit this (see Section 3.2 for an elaboration). Regarding trust, GDSs already have a reputation in the market, which they exploit through the creation of virtual agencies.

11 Travel agencies have also developed online booking systems, but these are still connected to GDSs. Travel agents are also using the Internet for information-seeking (and thereby bypassing GDSs). There are also web-based search engines dealing specifically with the lowest priced fares.
It is less likely that the possibility of direct contact between consumers and service suppliers will have an equally large effect on the industry. The customer is then faced with prohibitive search costs, ineffective bundling of the services required and trust deficiency. This is particularly true when it comes to small individual service suppliers in developing countries, and is not the case with large reputable airlines. The importance of trust generally increases with distance, particularly for differentiated goods. Small service providers do not have a reputation for reliability and will still need some kind of intermediary. The fragmented organisation of the supply chain is most likely to benefit a particular market segment, rather than spur overall growth in the market. Service providers will reach risk-loving, highly skilled (computer-literate) but low-income backpackers with small search costs (i.e. ample time to search). Highly skilled independent travellers with higher incomes and low search costs (i.e. enjoying searching on the Internet) also represent an increasing market potential although this market constitutes a tiny share of the overall tourist market.

More generally, online travel markets may be expected to have the greatest prospects in markets where:

- consumers have credible information about the quality of service providers, are protected by some system of warranties and are not averse to taking some risks
- money transactions are secure
- consumers and suppliers are computer-literate
- distribution costs are high
- technology is convenient (allowing e-tickets to be issued) and facilitates the creation of bundles of goods. For the supplier creating a bundle reduces price competition, but for the consumer it means greater variety. If consumers have preferences for particular bundles, they can trade off a higher price for a better match.

12 Sandelien (2003) provides an overview of the literature on the relationship between distance, trust and trade.
13 See, for instance, Goldstein and O’Connor (2000).
In both cases of online travel discussed above, however, the role of travel agents will be reduced. The potential exclusion of traditional intermediaries may reduce transaction costs, and if so, service suppliers may acquire a higher share of the value added. Being a small provider of tourist services, it is unlikely that Namibia will benefit greatly from this (although individuals with an established reputation may benefit). The ability to build trust and confidence could be even more important for information-intensive goods, partly since the cost of marketing on the Internet normally is low, making it more difficult to distinguish between high-quality firms and substandard companies just by their advertising. Goldstein and O’Connor (2000:28) claim that branding is important in order to develop e-commerce.

The effect of the Internet might be greater if competition between virtual agencies reduces the cost of being connected to a network and individual service providers co-operate. These issues are discussed in the subsections below.
3.2 The Internet may increase service providers’ participation in networks

In this section, I will analyse competition between GDSs. A GDS may offer a wider range of products than a CRS as it subsumes different CRSs. There are currently four major worldwide GDSs: Sabre, Amadeus, Galileo and Wordspan. Sabre is the largest, and is particularly important in the United States. In Europe, Amadeus is the largest, with a market share of approximately 60 per cent.\textsuperscript{14}

My main point of departure is that these distribution systems have market power and price their services in a prohibitive way from the perspective of an individual service provider in developing countries. Tourist businesses in developing countries are therefore generally not connected to such systems. Connecting a network is crucial for reaching a larger market. The question I raise is whether the Internet may change the competition between GDSs and thereby induce participation from firms in developing countries. There are at least three mechanisms at work. First, competition will increase from firms that sell their products online (see Section 3.1, above). There are two main groups of such firms. No-frills carriers such as Ryanair sell their services outside existing GDSs only. In addition, many reputable service providers are selling both through GDSs and directly to customers. The pressure from individual service providers may drive down booking fees.\textsuperscript{15} If not, GDSs will not be able to exploit their network externalities. A network externality is characterised by a situation where the value of participation of one user is increased when other users join and enlarge the network. If important service providers disconnect, GDSs will face negative externalities. Second, competition will increase due to lower switching costs across GDSs. Switching costs arise when an agent is locked in to a particular technology and it is costly to change to a new system, say a GDS. The Internet makes it less costly for consumers to change to a new system. Third, there may at the same time be a risk that the pressure on GDSs from important vendors undertaking direct marketing is passed on to service providers who do not have similar outside options (say small service providers in developing countries). GDSs are facing financial problems because airlines try to save on distribution costs by circumventing their services,

\textsuperscript{14} Galileo is the second most important with a market share of 22 per cent as measured in terms of the number of travel agencies having access to the system (see Buhalis, 1998; WTO, 1998b:8).

\textsuperscript{15} At an international travel conference in May 2003, representatives of GDSs acknowledged that fees will be reduced. http://www.travelmole.com
leading small GDSs to go out of business. Unless they diversify GDS services, there is the risk that some service providers will face higher fees. Individual service providers may also face tougher price competition due to the Internet. Below I will focus on the first two mechanisms, my point of departure being the literature on network externalities and switching costs (Katz and Shapiro, 1985, 1994; Farrell and Saloner, 1985, Shapiro and Varian, 1999; Economides, 1996).

With network externalities, consumers’ willingness to pay increases with the total number of units sold. A traditional GDS is characterised by three types of network externalities:

i) *Economies of scope*. The GDS system enhances the possibilities of exploiting the complementarities between goods and saving on distribution costs. The technical booking system is applicable to more than one good. The aggregate costs of providing a number of different goods are therefore lower than the sum of individual components. GDSs provide a number of different information goods such as information on prices of air tickets, hotels, cars, etc., each of which have low marginal costs. Bundling is an appropriate way of exploiting economies of scope due to technological complementarities in production, distribution or consumption (Bakos and Brynjolfsson, 2000).

ii) *Matching*. As more service providers are connected, more goods become available, facilitating the matching of a given number of consumers with service providers. As more service providers are connected, more customers will use the network.

iii) *Economies of scale*. The more service providers are connected, the lower the fixed cost per unit (average costs). The network is characterised by high fixed costs and low variable costs. To some extent fees reflect this cost structure. Service providers pay a (variable) fee per segment/ticket/booking dependent on their level of participation in the system. In addition there is a fixed cost for being connected to the network (particularly in terms of investments in specific technology). Higher participation may lower fees and thereby encourage entry.

Markets with network externalities are generally characterised by imperfect competition. This has also traditionally been the case.

16 Wordspan was sold recently because of financial problems.
with GDSs as revealed by the geographical segmentation of markets and concentration indices (see Box 1 for an overview of the anticompetitive mechanism applied by the GDS-owning firms). GDSs have a significantly higher market share in the country where the founder airlines operate than in other countries, and a few GDSs dominate the market (see above).\textsuperscript{17}

Box 1. GDSs and competition

The airline industry was among the earliest industries to explore the use of ICT through a privately-owned network of services. Electronic data interchange allowed the network participants to control, promote and sell their product globally. The network was called a computer reservation system (CRS) and refers to computerised systems containing information about carriers’ schedules, availability, fares and fare rules, from which reservations can be made and tickets issued (WTO, 1998b:4). Pemberton et al. (2001) distinguish between three phases in the development of the CRS and claim that the owning (airline) firms have kept their competitive advantage during all phases.

Initially only the big airlines developed internal seat inventory systems. Smaller air companies contracted out their seat inventory and reservation functions to carriers with CRSs. At this stage, owning carriers had a competitive advantage because of effective operational management of internal activities. During the second phase, CRS owners allowed other airlines to display information on their systems. The owners invested in CRSs while non-owning companies were charged booking fees. In addition to enjoying lower fees, owners also gained a competitive advantage by restricting fares and inventory from competing CRSs. One way of restricting fares was through the so-called screen display bias. When searching, systems were calibrated in such a way that parent company flights appeared before those of the competitor. Screen display bias was later prohibited by regulations, although some countries have made Most Favoured Nation exemptions for CRSs (see WTO, 1998b:10). The third phase of CRS development took place in the 1990s, when different CRSs were linked to each other through strategic alliances and partnerships and created global distribution systems (GDSs). Tour operators, travel agents, car companies, hotels and accommodation were also linked to these GDSs, making it easier for owners to control the whole value chain. During the third phase ‘control, dissemination and manipulation of CRS data played a vital role in sustaining competitive advantage for CRS-owning companies’ (Pemberton et al., 2001).

\textsuperscript{17} In the case of Amadeus, the three founder airlines Air France, Iberia and Lufthansa control 60 per cent of the shares and Amadeus’ market share in its founder countries is nearly 90 per cent.
The Internet represents a new phase in the development of GDSs, allowing consumers direct access through the Internet to a CRS or GDS. GDSs are adapting to this new reality by applying multi-channel distribution strategies. Sabre has introduced Travelocity. Galileo has acquired Trip Com. Amadeus has a stake in Opudo and E-travel. In addition to previous actors adapting their strategies, new actors with a background in communication services have entered the scene. Microsoft has, for instance, created Expedia, an online booking facility. Expedia and Travelocity are the largest online travel agents covering more than 50 per cent of the online trading market.

The new virtual agents mentioned above use existing platforms of GDSs (as Expedia is connected to the Wordspan system and Travelocity is connected to the Sabre system). For instance, Travelocity only provides information about service providers connected to the Sabre system. These ‘agents’ therefore continue the role of traditional travel agents, rather than substituting GDSs. They are dependent not only on the services provided by GDSs, but also on those provided by traditional travel agents. So far, they need travel agents to issue tickets and in many cases to take care of financial arrangements. Virtual travel agencies are therefore national in scope. More generally, travel agency services provided online are not very different from those provided by traditional ‘bricks and mortar’ travel agencies. On the demand side, customers can easily switch between virtual and traditional travel agencies. On the supply side, traditional travel agencies are able to turn into virtual agencies and vice versa. As for GDSs, it is easy for them to become virtual trade agencies.

To some extent, the close relationship between air companies and GDSs is being challenged, and the capital structure is changing. Sabre was separated from American Airlines in 2000 and Galileo was acquired in 2001 by Cendant Corporation, a vertically integrated multinational providing all types of travel services. GDSs become public, and discriminate less against non-owning carriers. Yet, the overall system has not become less concentrated.

The global distribution system has been accused of being discriminatory. The large virtual travel providers have met with the

---

18 The competitive assessment applied in this paragraph is in line with the approach taken by the European Community. See, for instance, The Commission of the European Communities. Case No COMP/M.2794-Amadeus/ggl/jv 21.05 2002:3

19 Avis, for instance, is owned by Cendant.
same accusation. The market is characterised by vertical restraints such as exclusive dealing, for instance American Airlines provided web-only fares to Travelocity, and not to Expedia, who did not agree to cut fees. Expedia on their side stopped selling Northwest Airlines for some weeks because the airline wanted lower fees. According to Forrester Research, Expedia tends to favour carriers with which it has marketing arrangements.20

Virtual travel agencies and traditional GDSs have the same type of network externalities. The same service providers are connected, the same bundling possibilities are available, but operational costs may be lower for virtual travel agencies. These network externalities are all mechanisms that have been available in the pre-Internet era, and it is an open question whether the virtual networks discussed above will influence them in any way. The main difference is that the consumer to some extent substitutes for the travel agent, does the matching and creates the bundle himself. To enhance consumers’ trust in all types of online travel (see Section 2.1), improved legislation is a prerequisite.

Individual service providers try to bypass GDSs by creating their own portals that are not connected to a GDS. Four American airline companies have for instance introduced Orbitz, and SAS has introduced online booking facilities which are not connected to any particular GDS. By creating Orbitz, the American airline companies could bypass traditional GDSs without being subject to mandatory participation in other distribution systems.21

For these new actors to be viable, bundling possibilities and trust seem essential. During the current Internet phase, consumers mainly package online on the web page of a virtual travel agency (build-your-own package) or combine individual home pages (say SAS and AVIS).

Individual service suppliers with a customer base and brand name represent a serious threat to GDSs, particularly when customers (as independent travellers) have low bundling costs or low preferences for bundling. In addition, the Internet has probably reduced the costs of running a GDS or virtual network. More important, the


21 Galileo claims that Orbitz is discriminatory.

http://www.cendant.com/media/pr/press_release.cgi/Travel+Distribution+Services/10982. In the United States, ‘mandatory participation’ has been required up to now. This means that airline companies owning a CRS should also participate in other CRSs. Through this participation, it is expected that airline companies make competitive offers at any CRS.
Internet has a greater potential impact on the switching costs between systems. It has also increased the compatibility between different online systems. Customers may switch from one online network to another with just a click, reducing the lock-in mechanisms characterised by previous GDS systems. The arrival of new virtual travel agencies, particularly agents unconnected with traditional GDSs, increases competition not because of a reduction in network externalities, but due to a reduction in switching costs. Each GDS has its own individual technological platform that requires specific investments in skill and capital (for instance in accounting and booking systems), and subsequently high switching costs for the travel agents. The Internet provides a more general technological platform and may therefore stimulate competition among the GDSs and between GDSs and individual service providers. If so, fees will decrease and the incentives to participate in networks increase.

At the same time, the main new market players are actually not new. As shown above, reputable firms are behind the new type of networks and GDS suppliers still control the market for electronic intermediaries. In markets with network externalities such reputations can be self-perpetuating, leading to increased concentration. The extent of vertical restraints also indicates that competition could be stiffer. In spite of these problems, it is reasonable to expect that the cost of connecting to a network will decrease and that will lead to increased participation both in traditional and new virtual distribution systems. Connection to a network is of the utmost importance for reaching a larger market and exploiting complementarities between goods. However, one also needs to bear in mind that consumers may gain through a reduction in prices due to lower distribution costs while individual service providers are faced with stiffer competition.

3.3 Destination marketing
UNCTAD devotes particular attention to one ‘virtual’ network and suggests that developing countries should create destination marketing organisations, co-ordinating individual initiatives. However, UNCTAD hardly discusses the requirements that need to be met if these organisations are to substitute for the travel agent, or whether governments should play a role. Proper standard-setting and cooperation between individual service providers and the government

---

22 See also WTO (1999).
are elements in such a strategy. Let me end this section with some comments on destination marketing in the light of the discussions in the previous sections. To facilitate a destination marketing portal, there are at least three reasons for government intervention:

▼ Direct network externalities. A portal has the same type of network externalities as those described in Section 3.2 (economies of scope, matching and economies of scale), but may be controlled by the individual service providers rather than by GDSs or their virtual substitutes. These externalities lead to co-ordination failures. An agent does not take into account that his connection to a network increases the value of the network.

▼ Public goods and information failures. The branding of a tourist destination, infrastructure and ‘wildlife’ are to some extent public goods and need government support. The same applies to the concept of trust, at least the type of trust that can be influenced by the government. In Section 3.1 the importance of trust was discussed. Without trust, people will not book online. Tourists do not have accurate information (see Section 2.1). The government should therefore control or guarantee the information provided through, for example, proper standard-setting, dispute resolution mechanisms and secure financial transaction systems. Otherwise there is the risk that ‘lemons’ tap the reputation of a particular destination.

▼ Complementarities and spillovers between pairs of tourist goods or across services (roads and air transport increase the value of accommodation). ICT investment facilitates communication and financial services facilitate online trading (see Section 2.2). Developing a booking system also requires the development of a telecommunication sector for communication and a financial sector for online trading. Tourism policy needs to develop complementary assets such as telecommunications, financial intermediation, trust and bundling facilities.

These market imperfections have led to market failures in the supply chain, and represent challenges for governments at any particular destination. Some parts of the tourist industry (GDSs in particular) are rather concentrated (see Section 3.2).

As discussed in Section 3.1, the travel agent is of particular importance in providing information and creating a package of goods (facilitating ‘one-stop shopping’). When the number of potential products is reduced (for example when there are a few service
providers from one country), it is easier to establish independent portals competing with the traditional intermediaries. As argued in Section 3.1, small individual service providers will generally not succeed with individual web-based solutions only, due to high search costs and a lack of high-profile reputation. Co-operation between firms through a network or portal is an alternative strategy. The establishment of a common portal for graded tourist resorts may reduce consumers’ search costs and lead to increased revenue for individual service providers. But the other side of the coin is that a new destination portal does not have existing customers and needs to build up a new clientele.

A new portal needs a booking mechanism for the individual service providers connected with it, links to information sites (so as to provide information), links to air companies (to facilitate the creation of packages), and some sort of arrangements for financial transactions, including electronic signatures and mechanisms for settling disputes. A portal needs to convey the message that its services are to be trusted. It also needs to connect to a number of search engines, otherwise consumers will not be connected.

4. Tourism in Namibia

The following section provides an overview of the tourist industry in Namibia based on secondary data as well as data from field trips to Namibia (in June and November 2001). Based on the theoretical approach in Section 3, the main purpose of this section is to illustrate some of the principal challenges faced by the tourist industry in one particular developing country, due to the rising importance of ICT. Namibia is of particular interest since its tourism potential is great. In addition, Namibia is a middle-income country with well-developed telecommunication and financial sectors, making it easier to exploit the opportunities created by the Internet than in many other developing countries without these complementary assets. To some extent, therefore, it serves as a test case for whether ‘the new economy’ makes a difference for tourism in developing countries.

I conducted 15 structured interviews with some of the main stakeholders in the tourist industry such as lodges, hotels and airline companies, tour operators, traditional travel agents and new ‘virtual’ agents. In addition, I interviewed representatives from the govern-

---

23 Air Namibia was the only airline willing to respond to my questionnaire.
ment and from the ICT industry. The interviews were not representative of the population and the selection of firms was also biased. I chose to interview the largest tour operators and the largest hotels because I expected them to be using ICT (through web pages, e-commerce or booking systems) more intensively than other agents in the sector and therefore more easily able to create link directly with customers. As a comparison, I interviewed some local community-based organisations, which served as network organisations for small tourist firms. I was particularly interested in responses to questions along the same lines as those analysed in Section 3.

▼ The current importance of travel agents and the structure of the supply chain. What roles are currently played by intermediaries, in relation to the size of the market and incentive structures between the actors involved?

▼ Whether the firms face any problems with their access to GDSs and how they have adopted ICT in general and as a marketing device in particular. Have firms developed e-commerce solutions and, if not, what kind of problems have agents experienced? How are service providers winning the trust of their customers?

▼ Private and public initiatives to increase co-operation, such as the creation of common platforms and of a destination marketing organisation.

Primary data will be presented in Sections 4.1-4.3 emphasising the importance of intermediaries, agents’ adoption of ICT and the government’s role in tourism and ICT policy. First, however, I will present some background material.

Tourism is the third most important sector in Namibia and its contribution to GDP is around 7 per cent. It is by far the most important commercial service exporter (92 per cent of Namibia’s commercial services are in tourism). Namibia had 861,000 tourist arrivals in 2001 and its expenditure on tourism was approximately US$ 300 million (in 1998). 24 It was one of a very few markets worldwide that grew in 2002, one of the industry’s toughest years in history. 25

25 Information provided from Galileo.
Namibia participates in the World Tourism Organization and in regional tourist organisations such as RETOSA (part of SADC). As a member of SADC, Namibia allows free establishment for SADC-registered companies in Namibia. During the GATS negotiation, Namibia is in fact the only SADC country without any limitations on foreign suppliers of tourism services.

Namibia’s market share in Africa is one-tenth of South Africa’s, the largest tourist destination in Africa. Tunisia and Morocco also have significantly higher market shares than Namibia. But Namibia scores higher than Tanzania (the most important tourist destination among LDCs). On average Namibia experienced a 12 per cent annual growth rate in tourism from 1995 to 1998 (WTO, 2000a). Although Namibia is a significant tourist destination in Africa, Africa accounts for only 4 per cent of the tourist arrivals worldwide and 2.3 per cent of international tourism receipts (WTO, 2000b).

There are several explanations for the growth of Namibia’s tourist industry. Increased consumer income in developed countries and an increased awareness of and preferences for new and exotic tourist destinations are relevant factors. Increased knowledge of foreign destinations and political stability in Namibia have spurred this growth. The tourist industry is characterised by product cycles where new tourist destinations regularly replace traditional destinations. We see that destinations in developing countries have started to compete with traditional tourist resorts located in the Mediterranean, particularly if these new destinations are perceived to be safer than traditional tourist destinations such as Turkey and Egypt. Namibia is an example of this. The growth is also a result of the globalisation process where service trade is increasing. Globalisation leads to an increase in business travel, which in turn leads to the expansion of leisure and personal tourism, which may have benefits for trade and investment. An increase in service trade increases tourism, particularly in sectors where the movement of individuals represents the main mode of supply. This is of particular importance for intra-regional tourists, such as business people from South Africa delivering services to Namibia.

---

26 For an overview of tourism in Africa, see Christie and Cromton (2001).

27 According to Denstadlie and Hjorthol (2002:73), overseas travel significantly increases with income, education and age (up to 66 years). Long-haul travelling has therefore increased during recent years. Other countries have similar findings (for the case of Germany, see, for instance, WTO, 2000b:1699).

28 Terrorist attacks, war and a severe acute respiratory syndrome (SARS) have changed this, particularly for countries in Northern Africa and Asia.
The market is divided into three different categories of international tourist. Visitors from South Africa (39.1 per cent), Angola (30.3 per cent) and Germany (10.5 per cent) are the most important (in terms of tourist arrivals). Tourist arrivals from countries outside Africa increased annually by approximately 20 per cent during the 1990s, which is higher than the overall growth rate (12 per cent). The growth rate has been highest for the richest income groups. The relative importance of intra-regional tourism is similar in South Africa and Namibia, and more important than in Kenya where inter-regional tourism plays a more significant role. The travel pattern and the reasons for travelling differ between these categories. Four out of five tourists travel for leisure, but the share of business travel is higher for African residents.

According to WTO (2000a), by 2020 more people will visit Namibia than the size of its population. WTO claims that the growth of tourism in Namibia will be higher than in the rest of Africa if Namibia maintains its conservation programmes.

4.1 Travel agents
For the big Namibian hotels, tour operators located in Namibia and Namibia’s wildlife resorts (covering all accommodation in national parks in Namibia), 60–80 per cent of bookings currently go through intermediaries, mainly located in the importing countries. For Air Namibia, the share is even higher. The commission rate varies between 9 per cent for the local air carrier and 35–40 per cent for the largest hotels and packages provided by local tour operators. These figures only relate to commission. If marketing costs and the costs of being connected to a GDS are included, distribution costs are quite high. For the local air carrier these constitute an additional 10 per cent. Distribution costs for the local service providers therefore vary between 20 per cent and 50 per cent of the consumer price.

4.2 GDSs, use of ICT and e-commerce
Apart from Air Namibia, the biggest tour operators and hotels, service providers were not connected to a GDS (Galileo and Amadeus in particular), mainly because their market base was too small to cover the costs. Instead, operators had connections with agents overseas. The variable fee is around US$ 5 per segment or booking. For room reservation this represents about 10 per cent of the
costs. The charge depends on the supplier’s level of participation in the GDS. So far, the establishment of virtual travel suppliers has not influenced the individual firm’s incentive (at least not in Namibia) to participate in networks, whether a GDS or an online global travel network. Generally, suppliers do not participate.

Apart from community-based tourist resorts located in remote areas without access to telecommunication facilities, and resorts in national parks, most firms interviewed had created their own home pages. Out of registered domain names (in Namibia under the .na name), around 20 per cent relate to tourism, but we do not know how many tourist sites are registered under the .com and under the .de (Germany) domain name. Many businesses put up web pages to protect their trademarks. Web hosting is cheap (less than US$ 100 a month).\(^{29}\)

Yet, few had online booking facilities and none had an e-commerce solution. E-commerce is a prerequisite for bypassing intermediaries. The combination of online (through e-mail) booking and payment confirmation by fax was widely used. The most important bottleneck for e-commerce was found to be insecure financial transactions. Although the same banks operate in Namibia and South Africa and e-commerce solutions are available in South Africa, such is not yet the case in Namibia. The financial sector is, however, implementing systems that may facilitate the development of e-commerce. Meanwhile, financial transactions are completed through intermediaries located in South Africa or by fax. It was also claimed that the country has not yet approved privacy laws and copyright protection, which are essential in order to secure the trust of individuals searching on the home pages of individual firms. In fact, firms could misuse their information by tracking individuals who have visited their home pages.

Namibia has not yet an approved ICT strategy or tourism policy (only draft policy documents). It is noteworthy that the current tourism policy does not address how ICT technology may facilitate an increase in the retention of revenue. When discussing the retention issue, the government focuses instead on ownership and leakage issues. The absence of electronic signatures, infrastructure security, privacy and data protection laws and online dispute resolution (for example, online handling of financial claims or complaints) impedes the development of e-commerce and the saving of distribution costs in tourism. Moreover, the question of increased confi-

\(^{29}\) I thank Ben Fuller for providing me with this information.
4.3 Portals and tourism policy

Few initiatives had been taken to create portals for the whole industry. At the time of my visit, the government had not implemented any measures to create a marketing destination organisation. A booking system for national parks has recently been created.\(^{30}\) Some firms, however, had been connected to new private travel portals such as [http://www.travelinnamibia.com/](http://www.travelinnamibia.com/).

Just as few government initiatives had been taken towards creating an ICT policy, such was also the case for Namibia’s tourism policy. Endless discussion, without any agreement on how the Tourist Board should be financed and on the appointment of a chairperson, led to a paralysed Board. Without any financial support, the Board could not initiate any activities such as marketing campaigns or standardisations of individual service providers.\(^{31}\)

Although the tourism policy recognises the government’s role in co-ordinating the private sector and the financial institutions, co-ordination of marketing strategies is hardly mentioned apart from the establishment of a National Tourist Board. Co-operation between the government and individual service providers has up to now been very difficult. The Tourism Board has now come on-stream, increasing these possibilities, although its US$ 1.5 million budget is extremely low. Its primary objective is marketing the country as a tourist destination. Branding Namibia has up to now been a very elaborate process. The Board has created a new logo and is in the process of introducing a whole new marketing strategy. An Act of Parliament made provision for tourist levies to finance the Board. It was originally proposed that levies would be raised through charges on business turnovers, but these are now to come through charges on tourists’ bills instead (15 per cent).

The tourism policy recognises the need to raise the level of its marketing expenditure, particularly in its key markets in Europe and in certain developed markets throughout the world, but barely discusses how the government can support marketing campaigns.

---

31 The Namibia Tourism Board seems recently to have come on stream.
5. Conclusion

Tourism is a service sector with untapped potential for developing countries. In this article I highlight the specific character of tourism as a bundle of complementary information goods. Tourism is an ICT-intensive sector with a spillover impact on other sectors. The main question addressed here is whether ICT, particularly the Internet, changes the traditional structure of the supply chain and thus opens up new possibilities for tourism suppliers in developing countries. Data from Namibia are used to illustrate some of the theoretical findings.

The Internet facilitates direct access to the customer. Reduction in distribution costs is of particular importance for the South. Today, intermediaries overseas certify and bundle the tourist product. I argue that without certification and the ability to bundle goods, it is unlikely that local providers of tourist services will be able to bypass the intermediaries located overseas without reducing the value of the product. While SAS may succeed with online direct marketing, a small provider in Namibia will probably fail. Accordingly, I find that the introduction of ICT in the tourist sector underscores the importance of market certifiers. Certification is needed, although new institutions may do it.

Direct marketing may lead to a reduction in the existing client base. At the same time, the Internet will stimulate groups of tourists with low search costs and who have low certification demands (such as backpackers or independent travellers) but they will not exploit complementarities between tourist products since bundling possibilities are weak.

Until new technological standards are available, I also dispute that new virtual travel agencies such as Expedia will increase competition between GDSs and thereby reduce the service provider’s costs of being connected to the network. Rather, these firms compete with travel agencies, not with GDSs since they generally use existing GDS booking engines and technological platforms.

Consumers’ switching costs will, however, reduce. Another GDS is just a click away. Competition will increase when new virtual agencies or reputable individual suppliers (such as airline companies) market online (outside existing GDS platforms). I conclude that fees will be reduced, making it more likely for service suppliers to connect with and reach a larger market. At the same time, the individual service providers will be faced with higher price competition.

In order to reach a larger market and to exploit complementari-
ties between products in the tourist bundle, I argue that individual service providers need to connect to a network. Currently, Namibian service providers are generally not connected to a GDS. The establishment of virtual travel portals in Namibia has not changed this yet.

Establishing a common marketing destination portal is a way to exploit network externalities and build trust, and thus, a mechanism to promote the tourist industry. In fact, the industry itself or representatives from the business community have initiated some portals, and the Namibian government is currently seeking to establish a marketing destination organisation. By creating a marketing destination organisation, one may explicitly stimulate co-operation between service providers. The introduction of ICT does not change the vendor’s individual gain from bypassing the travel agent. Tourism suppliers in Namibia therefore need to co-operate in order to achieve disintermediation. Such efforts may increase the tourism profit in the Internet age. Major structural changes of the industry are probably still yet to come.

References


UNCTAD, 2002b, *Background Paper on Developments and Main Issues in Electronic Commerce and Information and Communication Technology*, Geneva:
TD/B/COM.3/49. December.
World Tourist Organization, 1999, Marketing Tourism Destination Online: Strategies for the Information Age.
World Trade Organization, 1998a, Tourism Services: Background Note by the Secretariat, Geneva (S/C/W/51).
World Trade Organization, 1998b, Air Transport Services: Background Note by the Secretariat, Geneva (S/C/W/59).
Food Aid and Human Security
Edited by Edward Clay and Olav Stokke

The future role of food aid is in question. This matters because food aid has been historically a major element of development aid to support longer-term development and the primary response to help countries and people in crisis. Doubts about food aid are arising because there is a growing mismatch between the new circumstances produced by rapid political and economic change and the international arrangements and institutions for food aid that are predicated on an earlier reality.

In an environment of risks, uncertainty and rapid change, prevailing in the 1990s, food aid and other assistance have increasingly been organised as part of efforts to assure human security in terms of livelihoods, food, health, a sustainable environment, personal and political security. However, to what extent do is this multiplicity of goals realised in practice? To what extent the modalities and institutional arrangements for aid permit them to be realised? It is on institutional questions therefore, that this fresh examination of food aid focuses in particular.

"This is an important book, edited by two authorities involved in analysing food aid for many years, that should be read by those concerned with the future direction of food aid.

In many ways, this book represents an important watershed in the way food aid and its users are analysed, at least on this side of the Atlantic. Gone are the tired references to disincentive effects based on prejudice and anecdotal evidence. In place of calls for an end to food aid, and doubts about its future, a new and enlarged role is sought.

D. J. Shaw in Development Policy Review

Food Aid and Human Security
2000, 408 pp.
US$62/£45.00

www.frankcass.com
Debates

Mathieu versus de Soto: A Comment

Tor A. Benjaminsen and Espen Sjaastad

In the December 2002 issue of *Forum for Development Studies*, based on issues discussed at a seminar in Oslo in September last year, Paul Mathieu outlines some of his concerns regarding Hernando de Soto’s approach to poverty reduction through formalisation of property rights (Mathieu, 2002; de Soto, 2002a). These concerns, rather than constituting an aggressive attack on central components of de Soto’s approach, appear as reservations – aspects that should not be overlooked when a process of formalisation is undertaken in the context of rural Africa. Mathieu even begins his article by outlining what he believes to be a consensus between his own views and those of de Soto. In a sharp rejoinder, however, de Soto not only rejects this consensus but goes on to accuse Mathieu of academic blindness, paternalism, and – potentially – covert racism (de Soto, 2002b).

We believe de Soto’s theory is a powerful one, and we agree that formalisation of rights over assets and their integration into unified property systems can be an important tool in alleviating poverty. And some of the points de Soto makes in his reply to Mathieu are reasonable enough. For example, property is not just about social relations but also about how these relations manifest themselves with respect to valuable objects. Also, contrary to what Mathieu seems to imply, the complexity of rural tenures is not in and of itself a problem for formalisation; the intricacy of property regulations and contracts found within most Western nations is ample evidence of how well-designed property rights systems are capable of efficiently dealing with complexity.

Yet Mathieu also raises legitimate points about how and why de Soto’s approach to poverty reduction may stall or even be counterproductive in some African settings. We think these points are im-
important enough, and de Soto’s reply to them dismissive enough, to merit further comment.

At a national level, Mathieu supplies two simple assertions: for a formalisation process such as that envisaged by de Soto even to get off the ground, (1) the State must be willing to provide poor people with secure and formal rights over their assets, and (2) the State must be capable of providing such rights. These assertions seem self-evident, but the fact that they are obvious does not mean they can be neglected.

African governments – and the officials with which they are populated – may be unwilling to embark on a formalisation programme aimed at the poor simply because they stand to lose too much. For example, some of the most dynamic property markets in Africa are found in and around urban centres. As these expand, lands under customary tenure are converted into residential, industrial, and commercial areas. The increase in land asset values that attends this conversion is enormous, and the ones often best placed to benefit from this windfall are precisely those – central and local politicians and bureaucrats – that presumably would be charged with ratifying and implementing reforms. And in awarding secure rights to customary land holders, government officials may lose not only a source of wealth but control over the very process through which this wealth is generated.

Of equal concern, if willingness is indeed present, is the capacity of African states to implement and enforce wide-reaching and unified systems of property rights. Dr de Soto acknowledges that creation of such systems is costly, and that the modern systems held up as models required many years to evolve. The current context is one where African states, many of which are not only rife with corruption and mismanagement but also debt-laden and in a process of massive retrenchment, are expected to hurry the formalisation process along. And it is not as if the process involves one mighty initial effort whereafter everything takes care of itself; huge costs attend the maintenance and continuous refinement of such systems. The Institute for Liberty and Democracy, with de Soto at the helm, has achieved noteworthy success in formalisation of assets – and particularly with respect to small businesses – in Peru. But it remains to be seen whether such success can be replicated in an African context, and particularly in rural areas burdened with what is frequently both a flimsy and predatory state presence.

Though largely neglected in his earlier writings, de Soto’s insistence at the seminar last September that the formalisation process
should be demand-driven is welcome. However, both multilateral organisations and African governments have tended to take a rather different view in the past. At community level, de Soto (2002b: 379) also states that he has ‘never found a shantytown or a tribe in a developing country that does not want to have their assets protected by good property law’. Thus, a latent demand is basically always there and systems that do not meet local approval simply are not good enough; if you just find the right system, people will always want it.

But this ignores the real problem. Communities – be they shantytowns or tribes – rarely exhibit a single, unified, collective will with regard to these questions. One woman’s ideal or even acceptable system is not necessarily that of another, and the statement that ‘once you are in the field and aware of grassroots arrangements, all rights are compatible and can be clearly represented’ (de Soto, 2002b: 376) simply is not true, at least not universally true. A nephew may insist he is the owner of a plot that his uncle will insist has only been lent to him. According to ‘grass-roots’ arrangements, a single piece of property may have half a dozen owners. More generally, some rules and the rights that they generate – common examples are those concerning succession and alienation – operate at a collective level, and there will often be profound individual disagreement about their precise existing and future content. Claims may be advanced through appeals to widely disparate legal traditions, historical narratives and notions of identity.

Of course, such problems of collective action need to be sorted out at some point regardless of whether or when a formalisation process is implemented. But this brings us back to Mathieu’s point about process. Incompatible rights may, in the absence of formalisation, crystallise quite organically and painlessly. Formalisation, on the other hand, may in and of itself trigger a process of increasing opportunism, conflict and chaos. Ideal systems of property may indeed exist, but the process of getting there – of finding them and implementing them – may be very painful indeed. Western systems, good as they might be, were not won without much conflict and misery.

The point being that while some property systems may be ‘ready’ or ‘ripe’ for a formalisation process, others may not. Dr de Soto, on the origins of Western law, states that ‘it was born in the real world and bred by ordinary people through trial and error long before it got into the hands of academic commentators’ (de Soto, 2002b: 382). It was similarly born and bred, we can assume, before it got into
the hands of formalisers, be they Roman jurist consults or Peruvian economists. Yes, law has to be discovered before it can be studied; but for there to be any law to discover, it must first evolve.

It is an empirical fact that many instances of formalisation of property rights in Africa have led to more, rather than fewer, problems of dispossession, conflict and economic regression. These may have been ill-designed and badly implemented by governments of inadequate ability and insufficient commitment. But that is often how things go in Africa whenever big Western ideas are imported wholesale and put into action. If de Soto finds his ideas distorted or poorly understood by Dr Mathieu, he may be in for a shock when he sees what some African governments and international aid donors are capable of doing with them. Mathieu’s efforts to highlight some of the ways in which de Soto’s bandwagon might derail should, we believe, be welcomed. They are surely not deserving of the rubbing and personal attacks with which de Soto greets them.

Hernando de Soto has, since the publication of The Mystery of Capital (de Soto, 2000), travelled widely and spoken to numerous government leaders about formalisation of property rights for the poor. The Institute of Liberty and Democracy is now involved in projects or discussions with more than 20 governments about reforms; governments that, according to de Soto himself, have widely differing objectives with such reforms. How optimistic is de Soto about the commitment of the African state leaders he has met to go forth with changes that may enrich the poor at the short-term expense of at least some powerful individuals and groups? How convinced is he that reforms will be designed and implemented in a manner true to his intentions? Does he really believe that property formalisation can be implemented anywhere without significant loss to anyone? It will be interesting, in the coming years, to see whether what amounts to a grand experiment in fact achieves the admirable objectives invoked by its author.

References
An Additional Comment on de Soto

Paul Mathieu

A brief additional comment on de Soto: ‘Yes, but... – it is not so simple.’ Hernando de Soto proposes a fundamental and new message on the importance of (access to) information representing property for land tenure security, and – even more important – for the conversion of land assets (temporary or definitive) into capital.

His policy conclusion is: ‘Simple systems of titling will help the poor and stimulate investment and income generation.’

This is important, but it may miss some points. I want to add a simple but fundamental caveat to his (too) simple policy conclusion:

‘Yes, but other conditions are needed and it would not be simple in many cases.

There will be (big) winners and (big) losers in the process, certainly in most of rural Africa. Rent-seeking opportunities will be created in the transformation of tenure systems, and may ruin the virtuous circle envisaged.

Empowerment of the rural poor, improved access to scarce information and improved legal literacy of the intended beneficiaries are also needed.’
Possibilities of Global Governance in the Age of Globalisation

Henri Vogt

In an assessment of the effects of the World Summit on Sustainable Development (WSSD), held in Johannesburg from 26 August to 4 September 2002, a central question arises: to what extent can this kind of an event, directed from above, influence global developments and through what mechanisms? In other words, what are the chances of global governance being able to solve global problems and create a new willingness and ability in the world to work to this end? Or on a more general level: can political decisions direct the course of events in the globalised world, or does globalisation in fact lead to a new state of anarchy and an absence of control?

According to cynics, attempts to achieve global governance are bound to remain at the level of useless rhetoric. Such attempts are, at best, merely reactive, unable actively to guide the world’s development. Paradoxically, some cynics are also optimists: in their view, the risks to the world are exaggerated, and societies are always able to solve their problems in the course of time. There is also a normative aspect to be considered: it is inadvisable even to attempt to govern the world, as all such efforts could, in the end, restrict the individual’s freedom and society’s openness.

This pattern of criticism of global governance was very much evident at the WSSD. The Summit was considered insignificant because the issues to be discussed were all too difficult and complex – especially in a world where the United States is clearly intent on pursuing its own unipolar policies and does not fully contribute to multilateral international cooperation. Moreover, even the concept of sustainable development has been watered down and has thus fallen into the ever-increasing category of overused political slogans. So, according to some, the Summit was all empty words, mere theatre, unable to make any real positive impact on the future of our planet.

1 The crisis in Iraq has naturally made this problem even more acute.
2 In the Nordic countries, for example, this type of criticism was rather common. In Finland, an article by the Minister of Transport and Communications, Kimmo Sasi, in the daily Hufvudstadsbladet (17 Sept. 2002) was a typical example: ‘I
Another line of criticism at the WSSD was directed at the decisions made, especially at the paucity of both temporal and quantitative goals, and the fact that in many cases things did not really move forward from the decisions made at previous international events. In the opinion of many activists from non-governmental organisations, the Summit’s resolutions were in fact depressingly weak, and some of them even argued that the Summit should have been called ‘Rio-minus-ten’ rather than ‘Rio-plus-ten’. To give a random example, Barbara Stocking, the Director of Oxfam UK, made the following statement: ‘In terms of producing any real plan for tackling poverty in the long term, for example by addressing agricultural subsidies, the Summit has proved a disappointment’ (*Financial Times* 4 Sept. 2002).

This latter form of criticism was certainly justified in many respects, but it nevertheless overlooks the broadness of the WSSD and all the projects related to it. The Johannesburg Summit was indeed so broad that it was basically concerned with all aspects of life. Therefore it also had, and has, the potential to affect the state of the world in a multitude of ways and through many different channels. One can, to begin with, separate the decisions made from the process itself.

As regards the former, the actual results of the Summit can indeed be criticised for lacking tangibility and legal validity. On a positive note, however, the decisions (for example) to stabilise fishing stocks at the level of natural reproduction, or to ensure safe drinking water and sanitation for at least half of the two billion people lacking them at the moment, are crucially important with regard to sustainable development. At least as noteworthy as these kinds of decision listed in the Plan of Implementation, the central outcome

---

3 According to Paragraph 24 of the Plan of Implementation: ‘... we agree to halve, by the year 2015, the proportion of people who are unable to reach or to afford safe drinking water [...] and the proportion of people without sanitation.’ In Paragraph 30a, the Plan commits the signatories to implementing the following: ‘Maintain or restore [fishing] stocks to levels that can produce the maximum sustainable yield with the aim of achieving these goals for depleted stocks on an urgent basis and where possible not later than 2015.’
of the WSSD, are the different partnerships and initiatives introduced during the Summit – over 200 of them altogether. The range of actors varied greatly in both the partnership projects and the initiatives, and often both public and private actors were involved. The thematic spectrum of these projects was virtually all-encompassing: from supporting development in mountain areas to an educational initiative for sustainable urbanisation, from acknowledging biodiversity in the mining industry to the initiative for developing sustainable agriculture.\footnote{The official names of the partnerships and initiatives mentioned are: ‘Mountain Sustainable Development: International Partnership’, coordinated by the Swiss government; ‘Training Initiative on Sustainable Urbanisation’, coordinated by the United Nations Institute for Research and Training; ‘Partnership on Mining and Biodiversity’, involving the International Council on Mining and Metals and the World Conservation Union; ‘Sustainable Agriculture and Rural Development’, coordinated by the Food and Agriculture Organisation (FAO).} Even what one might see as unholy alliances were established, such as the partnership between Israel and Jordan for the protection of the Dead Sea.

But the Johannesburg Summit (and the whole series of multilateral conferences organised by the UN, as well as their side events) also has an influence, through a long-term process of gradually shaping human consciousness. The Johannesburg events were efficiently communicated worldwide through the media. The extent of the Summit’s visibility in the media, at least in Europe, was actually surprising, much more, for instance, than had been the case with the Monterrey Conference on development aid in the spring of 2002. The concept of sustainable development emerged once again at the centre of the world’s political consciousness – it may be that because of the tangibility of the concept, emphasising it time and again can truly affect the way people act in relation to their environment. Indeed, and especially since the Plan of Implementation of the WSSD is in many respects a compromise, it can be claimed that influencing as a process is essential for events such as the Johannesburg Summit – eventually more important than the actual decisions made.

It is not, however, particularly easy to estimate how strong this influence through a process is or can be. What kind of factors should be taken into consideration when we think about the significance of the process in the long run? At least three, partly interdependent factors or viewpoints are worth mentioning. The first of these is that of \textit{power}. The notion refers to the ability of the process to influence the universal distribution of power and affect the willing-


ness of people to lead things in a certain direction. In many ways the element of power is the most fundamental of the three. The second factor or viewpoint is mutual understanding between different actors, states, organisations and businesses. The term alludes to the possibilities of these actors to create common norms and dialogues as a result of the process – in more theoretical language, to create a common regime. The third factor is a kind of psychological state of mind that a political process may help to create. The notion of hope conveys the elementary aspect of this state in our analysis.

**Power**

Power is of course generally understood as power over somebody, as power to make decisions, as power to bring about action – by using force if necessary. Power in this conventional sense attracted much attention at the Johannesburg Summit: the delegates emphasised concrete action and implementation over fine-sounding but empty words. The primary question that arises when we think of power in this traditional sense is how legally binding multilaterally achieved decisions are, or how binding they are perceived to be. What is clear is that states do not necessarily have to ratify the agreements made in international conferences, or that they may not follow – either intentionally or unintentionally – the objectives they themselves have ratified. In the case, for example, of the international climate protection objectives, Finland – a country that basically tries to implement its international commitments as well as possible – has successfully carried out its commitments to cutting down on sulphurous anhydride emissions, whereas in the case of nitric oxides the country has not even come close to fulfilling its commitments. Developing countries, in turn, have not necessarily even had the capacity to estimate what kind of objectives they have in practice committed themselves to in international fora. It is not being excessively pessimistic to argue that when it comes to these kinds of difficulties of traditional power politics – difficulties of implementation and shortage of capacity – the effects of a process such as Johannesburg may be very limited indeed.

But power can also be understood in a much broader sense. The Johannesburg meeting can, first of all, be viewed as an attempt to seize power, that is, to change the distribution of power in the world. A historical parallel with the so-called first great transformation is illuminating in this respect. Karl Polanyi’s classic, *The Great Trans-
formation, written in 1944, is a presentation of how, in the early stages of industrialisation, the growth of economic activity called for a strong role for politics. Thus, politics, which gradually became more and more democratic and involved ever larger groups of people, necessarily developed so that at least some kind of social peace could prevail despite the problems created by rapid economic change. From this perspective, the multilateral UN summits are particular attempts to bring political power back alongside economic power, to bring politics onto the global stage too. Johannesburg could thus be thought of as representing some sort of Second Great Transformation (cf. Hettne, 2002; Teivainen, 2002).

Closely related to this, it can be asked whether an event like Johannesburg can also limit the power – most of all economic power – used in some other global arenas. Alex Callinicos (2002: 325), among others, has suggested that such international economic organisations as the Group of Seven (G7), the International Monetary Fund (IMF) and the World Trade Organisation (WTO) seem to have three central functions. First, they are a means by which the United States engages other wealthy industrialised countries in its own objectives. Second, they offer a forum in which the European Union, Japan and the United States can negotiate a way out of their conflicts. And third, they help these countries or regional blocks to put pressure on other countries, the great majority of states, to adopt the objectives of the dominant powers, usually according to the neo-liberal, so-called Washington consensus. One could argue, then, that since all the actors were present in Johannesburg, and able to express their viewpoints (it really was a multilateral event), the Summit presented an opposite force to the dominant way in which power is exercised worldwide. It should perhaps also be noted that learning to cooperate with new and different types of actors, above all the business community, could strengthen the whole United Nations system. We could, perhaps, even talk about the empowerment of the UN (cf. below). Indeed, Georg Kell, for example, from the office of the UN Secretary-General, stated at the WSSD that ‘at an institutional and political level you find a transformation in the UN as it learns to work with business and non-state actors. They were eye-openers here for us. It has energised the UN system’ (Financial Times 4 Sept. 2002).

Yet one can also contend that the Johannesburg process actual-

---

5 The return of politics and ‘the political’ into economic decision-making has, of course, been one of the central requirements of the critics of globalisation.
ly represented the complete acceptance of the current world order and that it ultimately contributed to its renewal. From this perspective, one possible analytical point of departure are the views of two post-Marxist political scientists, Michael Hardt and Antonio Negri; their book, *Empire*, published in 2000, has recently been the subject of a great deal of discussion (e.g. in *Millennium*, Vol. 31, No. 2). According to Hardt and Negri, the world is currently ruled by a limitless and undefined Empire that consists of a great many networks, civic organisations, states and businesses, which have all internalised certain sets of norms and ways of acting – to the extent that they are not even aware of these norms themselves. In this world of Empire, power is also part of an endless network, tangled in the conventions and rules of various human activities; power is both everywhere and nowhere. Understanding where power originates from or who it belongs to, and changing its balance, have thus become extremely pressing tasks.⁶

Although Hardt and Negri admit that despite its inherent problems the rule of the Empire is the best international system so far, their aim is to break its power. As this author understands their view, the way to do this is to generate some sort of new international solidarity, to develop a new pattern of global citizenship. It is difficult to imagine, however, what this new international solidarity would be like in practice, and what kind of a new state of affairs it would seek to build. But perhaps it would essentially be based on the willingness to seek mutual understanding – an issue that will be discussed in the next section of this article.

It is also noteworthy that the Empire is closely connected to the views we often express on globalisation – or maybe the Empire is globalisation. This is because of the dualistic nature of globalisation. On the one hand, it is a consciously governed and guided collection of phenomena, determined by shared international sets of norms and resolutions. On the other hand, it seems to be a virtually uncontrollable process: it has created a world where power is so decentralised that it cannot really be exercised. Globalisation thus seems to derive its energy, along with a great number of other characteristics, from the tension between governance and anarchy; and it represents, in the end, both.

Finally, power can also be viewed from the perspective of empowerment. Empowerment, and the closely related demand for

---

⁶ As I understand Hardt and Negri, their Empire is not too far away from such mythical notions as ‘international community’ or ‘world opinion’.
various capacity-building measures, has been one of the most central terms in development discussions over the past decades; recently it has probably been even more prominent in the debate. Empowerment is also emphasised throughout the Johannesburg Plan of Implementation, especially in relation to women, children and indigenous peoples, and the extent to which it is possible for them to put sustainable development into practice. Likewise, in the section dealing with the problems of Africa, capacity-building measures occupy a central role: the Plan states, among other things, that the participating states should ‘provide financial and technical assistance to strengthen the capacities of African countries, including institutional as well as human capacity’. The success of various capacity-building efforts is, above all, a question of conventional executive power, determining the extent to which resources are actually provided to those who need them. It is well worth noting, however, that people’s belief in their own capacity also plays a crucial part in this respect. Empowerment is thus also related to the notion of hope – the final theme of this article.

Mutual Understanding
Another parameter for the influence of the WSSD is the mutual understanding that it may produce among different actors. At the grass-roots level, an event like the Johannesburg Summit is, of course, all about people and the contacts between them; for many individual participants the Summit was undoubtedly a unique opportunity to meet people from different cultures and to find common areas of interests with them. As Kaarina Järventaus entitled her article in the Finnish daily, *Helsingin Sanomat*, ‘States Traded and People Met’ (5 Sept. 2002). Indeed, one can wonder to what extent Johannesburg and other similar worldwide get-togethers help in bringing about a new global but still individual citizenship, in generating new abilities to think globally, in creating global reflexivity as well as inviting a global sense of responsibility.

Above all, however, mutual understanding may develop through the creation of partnerships and initiatives of the kind mentioned

---

7 The quote is from the beginning of Paragraph 59a, although similar ideas are also presented in many other paragraphs in the section on Africa. The whole paragraph reads as follows: ‘Provide financial and technical assistance to strengthen the capacities of African countries, including institutional and human capacity, for effective disaster management, including observation and early warning systems, assessments, prevention, preparedness, response and recovery.’
earlier. Cooperation between different actors and the involvement of new actors are usually primary objectives of these partnership projects; according to the current rhetoric, cooperation between governments and civic organisations is particularly important. In this respect the parallel with the first great transformation would once again seem appropriate. During the great transformation in the 19th century, many new actors became part of decision-making procedures and thus also assumed political responsibility. In the long run the fact that many different interest groups came to bear the responsibility for the development of society also produced societal consensus and understanding, although a number of serious conflicts had first to be settled.

It is interesting, however, that the aim of involving all relevant actors in global governance processes is unprecedented in Western politics and political theory. While politics in the West have, throughout history, been based on exclusion, on building boundaries, in Johannesburg a totally different form of politics was pursued: a politics where boundaries between different actors had become blurred; a completely inclusive form of politics (cf. Ojakangas, 2002.) This idea or ambition of all-inclusiveness is, in fact, one of the defining features of the era of globalisation. There are, however, two major risks in this idea. First, conflict in politics means that people are heavily involved. But if this conflictual component disappeared and people became indifferent to public affairs, politics would begin to resemble bureaucratic decision-making with no real democratic control. In this sense, creating mutual understanding should not be equivalent to eradicating conflict. Second, all-inclusiveness is bound to remain an illusion, but if we nevertheless believe that we have achieved it, we may easily forget those who are still excluded, the lowest and most alienated strata of society.

8 In his opening speech Nitin Desai, the Secretary-General of the Johannesburg Summit, described the nature of partnerships accurately: ‘Partnerships come in basically to connect the dynamism that we see at the local level with the commitments which the governments need to make. We need both. Not one or the other. Both. Partnerships without the commitments of governments will not work.’ http://www.johannesburgsummit.org/html/documents/statements/2608_desai_opening_speech.pdf ; site visited on 18/10/2002.

9 For example, Mika Ojakangas (2002: 18) formulates this point as follows: ‘The basis for a political community has so far been excluding others. It is namely because of this [...] that we now need to find forms of interaction beyond conventional politics. In other words we must try to answer the question, what would such a community be that would not be based on exclusion but that would nevertheless be political.’
The partnerships and initiatives agreed upon in the WSSD often spring from the idea of dialogue. This naturally reflects the recent development discourse where the demand for new forms of global dialogue has been expressed time and again. Apart from the fact that as many actors as possible should take part in these constructive dialogues, the overarching idea is that they be always based on some kind of mutual trust and respect, on mutual accountability. For example, one central objective of the initiative ‘New Partnership for Africa’s Development’ (NEPAD), introduced by a number of African countries in 2001, is to create possibilities for an equality-based dialogue with the rest of the world and especially with the wealthy Western countries. To achieve this is mainly a question of attitude: both African countries and the former colonial powers should finally move beyond the burden of colonialism and begin to act together, respecting each others’ traditions and seeking mutual benefits. How to make this happen in real life, is, of course, much more difficult, as the parties in this continental dialogue have such different resources – a problem that so often arises in world politics. In any case, it is important that new forms of cooperation, such as NEPAD, are supported in the world’s political arenas; even the possibility of bringing about constructive dialogue deserves all the support it can get.

If one wishes to consider the issue of mutual understanding from a more theoretical point of view, the concept of regime provides a useful starting point. In other words, we need to ask how important a summit like the WSSD is or can be in terms of creating commonly binding principles, patterns of thought and conventions. In the light of the speeches given in Johannesburg, it seems that a regime of sustainable development is in many respects strong and well established. The participating groups had basically similar views on

---

10 The NEPAD initiative is given special support in the Johannesburg Plan of Implementation.

11 The broadest new world-level initiative that aims at mutual understanding is undoubtedly the ‘Global Deal’, a cooperative deal that UN Secretary-General Kofi Annan presented in Monterrey in the spring of 2002. He made clear that the Deal would mainly be about cooperation between the North and the South and would be based on the common understanding that the aid of rich countries to the poor is not about good will or pity but about mutual benefit.

12 Stephen Krasner (1985: 4), one of the founding fathers of regime theory, defines the concept of regime as follows: ‘Regimes are principles, norms, rules, and decision-making procedures around which actors’ expectations converge. [...] For instance, a liberal international regime for trade is based on a set of neoclassical economic principles that demonstrate that global utility is maximised by the free flow of goods.’
how to identify general worldwide problems and their possible solutions. Moreover, all participants strongly emphasised the importance of concentrating explicitly on action; ‘we have to move from word to deed – now’ was the common message. No wonder, then, that Bartholomäus Grill, a journalist on the German newspaper Die Zeit (5 Sept. 2002), could start his summarising article on the Summit as follows: ‘To fight against the poverty of billions of people. To restore natural resources. To change our production and consumption habits. Who is the one talking? A Greenpeace activist? An opponent of globalisation? No. It is Ian Johnson, Vice-President of the World Bank.’ It should also be noted that Johannesburg – and the whole concept of sustainable development – is an attempt to unite different smaller regimes – the climatic regime, the desertification regime, the poverty alleviation regime – into a kind of super-regime.

The concept of regime implies the idea of continuity, even stability; a regime is always based on inertia. In its role as a sequel to the Stockholm Summit of 1972 and the Rio de Janeiro Summit of 1992, the Johannesburg Summit was part of a historical continuum. The question then arises whether merely investing in low-level implementation can uphold the continuity and momentum of the idea of sustainable development, or whether this continuity actually requires ritualistic meetings of the whole world, such as the WSSD. Another significant feature of a regime is its ability to resist the power politics and vested interests of single concentrations of power; in that sense a regime is also a safety net. To what extent the regime of sustainable development can gradually oppose for example the EU’s willingness to limit free trade according to its own agricultural policy interests, is naturally impossible to estimate. But whatever the case, at least the regime can try to make people aware of the existence of this kind of problem.

Hope

Mentioning the notion of hope in this context may at first seem strange, but it actually has great significance in evaluating a process such as the WSSD. The concept is used in relation to the faith of individuals and communities in being able to improve their own circumstances. In other words, what is significant is the level of people’s psychological awareness and their readiness to act, their willingness to work to change the way things are; hope can open up new horizons for the future. Whereas the two previous factors
primarily concern the official politics of institutions and organisations, hope is about the relationship of individuals – also as members of communities – with their living environment, ultimately with their own lives.

Empirical evaluation of the significance of hope is, of course, tricky, but a parallel with the events that brought about the collapse of socialism in Eastern Europe gives an idea of what the term means in this context. As is well known, the Helsinki Process, and especially the 1975 Helsinki Final Act, was an important factor in breaking down the communist systems. The process built a moral foundation, on the basis of which the Eastern European democratic dissidents could demand of communist political leaders that they account for their actions. It became possible to make states responsible for their deeds, and this seemed to give individuals new possibilities for action, new hope. Another crucial factor in dismantling communism was Mikhail Gorbachev’s assumption of power in the former Soviet Union in 1985, and the beginning of the policies of glasnost and perestroika. Above all, these policies made it clear to the masses that even communist systems could be changed. Hope, and the belief in the possibility of an alternative future, eventually led to the destruction of the communist systems (Vogt, 2000).

The question now is whether a process like Johannesburg can open up new horizons of hope and a sense of future, horizons that could make people fight for better circumstances for themselves and for the ones close to them – or even to act critically with regard to the structures of the global system. For example, to what extent can this kind of an event convince an individual of the importance of buying biodynamic products? Or products with a ‘fair trade’ label? In this respect, it is a fact that as a result of the WSSD the concept of sustainable development was once again promoted all over the world. The concept is based on a firm belief in the future, on seeing the future as open, unpredictable and full of new resources. Thus the message that the notion conveys is primarily optimistic: there is really a lot that can be done to make the world a better place to live in.

13 Interestingly enough, one of the most steadfast critics of globalisation, Naomi Klein, tells about such new horizons opening up in the globalisation-critical ‘counter-summits’: ‘The first time I participated in one of these counter-summits, I remember having the feeling that some sort of political portal was opening up – a gateway, a window, “a crack in history” [...]. This opening was a sense of possibility, a blast of fresh air’ (The Guardian Weekly, 17–23 Oct., 2002; the text is an excerpt from Klein’s book Fences and Windows).
This optimistic tone regarding the notion of sustainable development was also evident in the speeches and comments made in the course of the WSSD process. It was emphasised that sustainable development and the procedures it requires should be seen as an opportunity rather than a burden, including for the world of business. This is the same point that UN Secretary-General Kofi Annan made in a speech in the spring of 2002, when he stated that ‘far from being a burden, sustainable development is an exceptional opportunity – economically, to build markets and create jobs; socially, to bring people in from the margins; and politically, to reduce tensions over resources that could lead to violence and to give every man and woman a voice, and a choice, in deciding their own future’. 14

It should also be clear that where mutual understanding was found at the Summit – through different partnerships, initiatives and regime-building – this in principle creates new hope, hope for better dialogue between different actors and for the emergence of common views regarding a number of central problems and their solutions. It should not be forgotten either that even though many were disappointed with the progress made in Johannesburg – some members of civil society organisations must have felt more despair than hope – in the end a mutual agreement was achieved; multilateralism functioned at least in some way. Remaining without any agreement would no doubt have violated hopes of the globe’s ability to care about its own future.

All in all, when we think about the project of sustainable development, creating and maintaining hope seem essential. We can ask whether an African ghetto, seemingly doomed to absolute poverty, would survive, if it were not for the hope of a better day. The fight against HIV/AIDS is mostly about giving hope to the patients. Similarly, in the case of the above-mentioned NEPAD initiative, hope plays a crucial role. As the initiative is still merely an idea, it is particularly important that it produces among ordinary Africans a positive consciousness and a conviction in the possibility of change in relation to their own African opportunities. To what extent the Johannesburg process can really influence social problems and attitudes is, naturally, impossible to estimate. The question is often about whether people get the necessary first encouragement, or see a ray of hope – as they did in Eastern Europe – which inspires them to

act in order to make things better. The willingness to bring about change has to come from the grass-roots level, from people themselves.

Let us finally note that in many speeches in Johannesburg, the importance of hope was repeatedly alluded to, particularly hope for a happier future for generations to come. For example, the President of the European Commission, Romano Prodi, stated that ‘the citizens of the world look to us for answers. It is our duty not to disappoint them. Collectively, we have to show them we can harness the power of globalisation, give hope to the world’s poor and preserve the resources and the beauty of our planet’. In turn the British Prime Minister, Tony Blair, said, with a somewhat neo-colonialist undertone: ‘Yesterday, in some of the poorest parts of Mozambique, I saw children every bit as bright as children in affluent Britain. Full of potential. Full, despite all the challenges, of hope. But their life chances stunted by poor health, poor housing, poor education, poor sanitation.’ (Both speeches given at the WSSD on 2 Sept. 2002.) Words of this kind are, of course, primarily rhetoric; they nevertheless indicate that even global governance efforts must derive their energy from the level of individuals and their attitudes.

In Conclusion
It has certainly become obvious how difficult it is to assess the extent to which multilateral and international decision-making and governance can shape the course of events in our world. The chances of leading the planet in a sustainable direction through common political decisions are no doubt limited, but despite this the processes of global governance can change the world through a number of different mechanisms, even in ways that we are not necessarily conscious of. What is also important is that any particular change or step in the development process is always a result of the combined influence of many different factors and actors. For example, the processes of global governance might mean little or nothing without the involvement of the media. This multiplicity of channels of influence is, in fact, one of the characteristics of the world today, one of the main features of the era of globalisation. It was also the most conspicuous feature of the Johannesburg Summit.

Be that as it may, it is clear that too few concrete decisions were made at the World Summit on Sustainable Development, too few temporal and quantitative objectives were agreed upon. There was,
for example, no talk of alternative financial sources for sustainable development – I doubt whether anyone actually mentioned that even a fraction of the world’s military expenditure would be enough to reduce poverty considerably. But this limited scope in relation to alternative strategies is perhaps the price we have to pay if we want world politics to be based on multilateralism and equal distribution of power – to the extent that this is possible – in the future too. Indeed, preventing an excessive concentration of power is the primary function of all multilateral processes.

It remains to be seen whether this is just more wishful thinking as a result of recent developments at the level of world politics. As I finish making the last corrections to this article, the war in Iraq has been going on for a week.

References


This article is in three parts.¹ The first is a brief presentation of the research agendas drawn up by African participants in a project on urban research in the developing world called the Global Urban Research Initiative. In the second I discuss a dominant mode of donor financing of African urban research – the research consultancy – arguing that it has negative implications for the building of research capacity and for quality in research. The third part is a discussion of what we as researchers can do to influence aid agencies and other funding bodies to be more constructive in the way they channel resources to social research and in their use of researchers and their work.

¹ This article is based on a paper presented at the seminar ‘Beyond the neo-liberal consensus on urban development: Other voices from Europe and the South’, held in Paris, 15–17 May 2003, by N-AERUS (Network-Association of European Researchers on Urbanisation in the South).
with hundreds – sometimes more than a thousand – entries. GURI has resulted in a series of books and published articles. The following reflections are primarily based on the two volumes devoted to Africa (Stren, 1994; Swilling, 1997) and on various contributions in two thematic collections (McCarney, 1996; Stren and Bell, 1995).

Here are some of the principal observations regarding knowledge gaps and proposals for research agendas presented by the African contributors. First, a few words about their overall diagnosis. There is a general consensus that African cities are in crisis, and that the crisis consists of failing services and inadequate local government structures, shortage of housing and jobs, severe environmental problems, widespread poverty and increasing inequalities. It is interesting to note, however, that while African governments, civil servants and observers such as journalists normally attribute the urban crisis to explosive urban growth and adverse economic circumstances, the GURI analysts tend to see it as a result of failures in government. Since independence, states have failed to provide institutional and legal frameworks for the overall development of cities. Instead, individuals and firms are exposed to obstructionist legal norms, corrupt civil servants and pervasive informality.

Some important knowledge gaps were identified, particularly concerning the urban economy. Mohammed Halfani (1997) observes that there are no serious studies on the productive capacity and competitiveness of African cities, in relation to the twin processes of marginalisation and globalisation. On poverty and inequality, he points out how the current understanding is rather incoherent. Some statistical information exists, but is generally highly aggregated and mostly out of date, it refers to only a handful of cities and, even at that level, is rarely comprehensive. How poverty is gendered and how it affects people differently at various stages of the life cycle are important questions in international poverty research, and one might add, central to formulating policies for poverty alleviation. But these are matters about which we know very little from African cities. On the urban informal economy there is a remarkable amount of generalisation based on fairly thin ground. Such generalisations are often done on the basis of scattered case studies, whereas for a clearer picture very different types of data would be needed.

Even when the local (or international) research community may have valid and potentially useful information, it is rarely used. The Nigerian geographer, Akin Mabogunje, the doyen of African urban studies, observes that the relationship between researchers – ur-
ban or otherwise – and policy-makers is far from cordial or close in most African countries. Many decisions affecting urban development or mere day-to-day management of cities are taken without appropriate information, which might have been provided if the two groups had co-operated. This lack of co-operation means there are no well-articulated national urban policies that would have facilitated the identification of national priorities in urban development. But the approach of most African governments to the problems of the city tends to be piecemeal and project-oriented (Mabogunje, 1994: 38).

Mabogunje links both the poor relations between researchers and (urban) governments and the precarious state of urban research in most countries to the lack of articulated urban policies and to African governments’ dependence, intellectual and financial, on foreign donors. Urban research topics are determined more by the project objectives of foreign donors rather than formulated in a dialogue between local researchers and policy coalitions. African governments’ fragmentary approach to urban problems is quickly shown to be inadequate when confronted with the well-prepared proposals of the donor agencies:

The vacuum created is then filled by the research agenda and programmes of the donor agencies. However, since the basic preoccupations of these agendas and programmes reflect no more than the current state of the internal dynamics and power plays within the agencies themselves, the life of a particular research problem and its long-term and sustained investigation always remain very uncertain. The result gives the impression that the research preoccupations in African countries are no more than a reflection of current fads among donor agencies (ibid.: 38).

In some contrast to this severe indictment, the various African participants in the GURI exercise have a clear view of the place of research. On the whole, the various contributions from Africa’s sub-regions are remarkably sober and modest in the way they formulate the role of research in urban development. Here are few echoes of the 1970s, where research was alternately praised as a tool of liberation or condemned an instrument of oppression. Mabogunje sums up the role of research in this way:

…research must focus on providing the knowledge and information that will guide formulation of realistic policies and decisions to improve the situation in urban centres (ibid.: 42).
Mohammed Halfani, writing about East Africa, formulates a programme that may seem a bit more ambitious or even unorthodox when he states that:

formulation of an agenda for urban research…has to be a collective and participatory process involving all the actors engaged in the determination of urban development. In an ideal situation, this includes urban residents in their various categories, policy-makers, sponsors and scholars. In initiating such a process, however, it is the responsibility of scholars to prepare the groundwork for discussions, negotiations and compromises… An agenda for urban research must take into account the gaps in our knowledge of the process of urban development up to this point (Halfani, 1994: 54).

The research agendas drawn up by scholars from the various African sub-regions are remarkably similar: They all include urban poverty, urban management and governance, the urban environment and urban demography. Land use, shelter and the role of women in urban development are also mentioned by most. There is also an agreement that a large part of existing and potentially useful research-based information remains disjointed and has little chance of becoming incorporated in the governance process. Another area of consensus is that African urban research needs improved publication and dissemination systems. Researchers need better working conditions and resources for networking. All this is of course partly a question of money, which leads us to the second challenge of current African urban studies.

**Funding Research through Consultancies**

In the first two decades of independence, support for research in Africa came primarily from national governments, and research was carried out in universities and public research institutes. For a number of reasons, of which financial constraints are but one, institutions of research and higher education declined during the 1980s and 1990s (Sawyerr, 2002). Funds for research dwindled, and both natural science research and social research – urban and otherwise – are now overwhelmingly financed from abroad, by donors. Some organisations, such as the Ford Foundation, SAREC (Sweden) and IDRC (Canada) provide not only funding for specific projects or programmes but institutional support for research. These institutions also have strengthening of research capacities in partner countries as one of their objectives. There is reason to believe
that research agendas at least are drawn up in consultation with local scholars.

In the same period as international donor agencies rather than national institutions are increasingly financing university-based research, two other forms of financing applied social research through donors have emerged: the research NGO and the research consultancy. Urban NGOs with research or advisory services on their agenda are numerous in South Africa. Outside this country, the principal NGO-based urban research centres are ENDA in Dakar, CASSAD in Nigeria and the Mazingira Institute in Nairobi. All are fairly modest in size, and outside South Africa, research consultancies are probably the most important mode of financing urban studies. These are individual studies commissioned by donor agencies, which need research-based information as background to their various activities, and for planning, implementing and evaluating programmes and projects. In an overview article on the first five years of GURI, Richard Stren (1966), the project’s coordinator, discusses the chief characteristics of research consultancies and some of the unfortunate side effects of this mode of financing research. Such commissioned studies are rarely long-term programmes where it is possible to build up the longitudinal sets of data required for establishing the necessary knowledge base about current urban processes. Rather, they are usually very specified or narrowly defined, time-bound investigations, constrained by limited terms of reference. A report is required within a short period of time; however, the report is rarely published but more often subsumed in larger project documents written by expatriate teams of consultants.

African academics accept these terms because alternative funds are not usually available and they need to supplement their salaries. But the research consultancy as a mode of financing urban studies, in common with social research more generally, has a series of negative effects. First, it contributes to what may be called the pulverisation of research ideas and of the research communities. Research ideas are not emerging from a combined process of theorising, looking at available information and local definitions of knowledge needs. Rather, specific questions to be answered are formulated by the agencies and put before the researcher/consultant. Researchers have to compete for contracts; thus, research consultancies tend to undermine the perception and practice of research as a collective enterprise. Second, the selection of research topics is normally done by the agencies that are financing the work.
These agencies are driven by their own agendas or sometimes shared international agendas, not by an assessment of local knowledge needs. Third, since consultancy-based research is rarely published, it does not contribute much to training, and long-term capacity-building needs are overlooked (Stren, 1996: 115).

The result is a bulky and elusive ‘grey’ literature, and studies risk being forgotten by the agencies that commission them even before they are finished. Here is one illustration of this state of affairs, reported by Stren (1996: 113). In the early 1990s, the French Ministry for Housing and Transport invited the sociologist Isabelle Milbert to look into what could be done to make this type of literature more accessible. Milbert (1992) found over 300 major urban studies commissioned by a number of French government agencies during the 1980s. They were located in 14 documentation centres and most of them were never published or even circulated. The agencies that had commissioned these studies were rarely concerned with diffusion and, in many cases, their mandates had changed by the time the reports were submitted, further reducing the incentive to publish.

On Quality

In addition to concerns about pulverisation of research ideas and negative consequences for the building of research capacity through research consultancies, there is also reason to look into how this mode of financing affects quality in research. Consultancy reports commissioned by aid agencies are of very varied quality, sometimes shockingly low.

In my experience, donor agencies that commission evaluations, background studies or feasibility studies on whatever sector-related or project-related knowledge they have decided they need, do not look for quality, nor do they honour quality. In the case of reports commissioned from consultants, be it researchers or consultancy firms, the impression gained is that it is more important for the report to be delivered on time than for it to contain valid and reliable information and sound analysis. Since it is rarely published or even made available to the consultants’ peers, a report is not likely to be subject to any professional quality assessment. One often gets the impression that the whole exercise is a ritual rather than a search for answers, particularly when the urgency surrounding the commissioning is not reflected in any follow-up action (Vaa, 1995).

In some respects, quality criteria for consultancies overlap with...
those for research; in others they are rather different. Arjun Appadurai (1999:234) offers what he calls a naive definition of research, namely the systematic pursuit of the not-yet-known, based on a reasonably clear grasp of relevant prior knowledge, and following established protocols. In addition, the findings should be deemed interesting by a vocational and specialised community of assessors. Obviously the type of research Appadurai has in mind here is researcher-initiated studies funded on the basis of scientific merit, not practical relevance or usefulness. This is very far from the production of knowledge through studies commissioned by public or private bureaucracies where the professed aim is to establish a knowledge base for informed action. Aid agencies may with some justification argue that when they invite academics to do studies for them, they do not perceive this as channelling resources into research, or even as contributions to exercises in knowledge production, but simply as convenient ways of collecting information they do not have in-house capacity to collect themselves. It would be neither fair nor particularly useful to apply the same quality criteria to this type of studies as to academic research. However, if a study is worth doing, presumably it is worth doing well. My submission is that in a number of respects, quality criteria are the same for both types of work. Here are some of them:

- **Written reports should be readable.** In practice, consultancy reports are often under-edited, long-winded and cumbersome to read.
- **The information they contain should be accurate, comprehensive and relevant to the problem at hand.** In practice, reports are often quite selective in their presentation of background information. The quality of the data presented is rarely discussed and it is not unusual to find inconsistencies in different accounts of the same phenomenon, be it a project, a programme, a sector, an institution, a country or parts of a country. Logically, this means that either one or both of the accounts are factually wrong.
- **Reports should convey a sound grasp of the local situation and local institutional structures.** However, consultancy reports sometimes reveal limited knowledge and experience of local contexts and institutions. In fact, international consultants tend to reproduce similar analysis and proposals regardless of which country they are in, probably because their knowledge of
each location is so limited. This is particularly striking in poverty assessments (Hanmer et al., 1999). 2

There should be a clear link between the information provided, the analytical part of the report and the conclusions or recommendations offered. This is perhaps the quality requirement most often sinned against. Surprisingly often, recommendations seem to be mere add-ons with no apparent basis in the information that is presented and analysed (Diallo and Vaa, 2000; Whitehead and Lockwood, 1999).

These are minimum requirements. In addition, there are of course subject-specific requirements of expertise for producing knowledge in any chosen field. It should be self-evident that a solid professional background and a sound knowledge of the country and sector in question is a necessary condition for doing a competent job. These are perhaps the most frequently ignored requirements, but they invite a quite different discussion, so I will let them lie.

I am not saying that most consultancy reports suffer from these shortcomings, nor do I want to be more specific about the proportion that is below par. No doubt, a lot of excellent work is being done too. Let us not forget that Amartya Sen’s brilliant essay on poverty and famines started out as a consultancy report for the ILO (Sen, 1981). My point is simply that quality is variable and that the mechanisms to ensure quality are weak, if they exist at all. This should be recognised as a problem by aid agencies, consultancy firms and the academic community.

One plausible reason for the present state of affairs can be found in how the agencies that commission consultancies manage time. Most students and practitioners of development agree that development is a slow process, and that the financial resources for bringing about development are far from adequate, be it aid flows or developing countries’ own resources. But donor agencies have a tendency to manage and mismanage time in such a way that time has replaced money as the scarcest resource. Shortage of time has become institutionalised. This is also evident in the way studies are commissioned, time is rarely set aside for a proper discussion of the terms of reference, and the time frame for background study and fieldwork is often inadequate in relation to the issues to be covered. So

2 At a workshop on Aid and Academia, held at Uppsala University in October 2001, Raymond Apthorpe of Sydney University characterised World Bank poverty assessments as an intellectual crime against humanity!
is the time for writing, which may be part of the reason why consultancy reports are often so under-edited and cumbersome, and very rarely enjoyable or even interesting to read.

What is also striking is that studies that were of the outmost urgency when they were commissioned often get forgotten almost as soon as they are produced. Once an intervention is decided upon, aid agencies become preoccupied with speedy implementation and measurable results. But haste and urgency have in some instances peculiar parallels in non-action. The processes through which agencies decide which problems to deal with, and which to ignore, would be a fascinating subject for research; so would the processes through which a topic turns from urgent to dead overnight.

**What can be done?**

One step towards increasing the quality and hence the usefulness of consultancy reports would thus be for the agencies to be more reflective when they commission studies, and to allow time for more discussion internally and with the people who are going to do the study. Representatives from partner countries should also be involved in discussions on why a particular constellation of information, data and knowledge is needed at a particular juncture in time. This requires not only increased levels of rationality in the agencies but also a willingness on the part of potential consultants to ask possibly embarrassing questions about why a study is to be undertaken, instead of just accepting a given commission. One simple measure towards increasing rationality and reflexivity in agencies would be for them to promise feedback on studies they commission. This would also be a learning experience for all parties.

Bureaucrats may sometimes have unrealistic and diffuse expectations of researchers: that they will somehow be able to provide clear and simple answers to questions that nobody has been able to formulate. Or expectations may be so low that only the most pedestrian tasks are put before the researcher. Neither set of attitudes is likely to invite independent inquiry or foster the pursuit of quality in research.

Another road to improved quality in research consultancies, again involving both agencies and researchers, would be to look at procedures for increased quality in other fields of knowledge production and see if there are institutionalised practices that can be transferred. In the research community, quality is ideally assured both through set procedures and institutionalised communication.
between researchers. One set procedure is professional assessment of research proposals; another is peer evaluation before publication of reports. Institutionalised communication between researchers may take the form of reviews of published reports (books, articles) and a continuous diffusion and discussion of findings in encounters between researchers and between teachers and students. Researchers may often have scathing comments about the quality of each other’s work, but at least quality is a fairly constant preoccupation.

Consultancy reports for development agencies are not always made public, and they are definitely not subject to institutionalised peer review. There is rarely any professional quality assessment or broker between the team or firm that authors a report and the agency that commissions it. Perhaps agencies should employ expert readers, in the way publishing houses do, to assess quality in reports. Other procedures that might be borrowed from the academic world would be to insist that reports should be published and diffused, and to invite journals to review reports.

Consulting firms and individuals would soon learn both that quality pays and that to continue to ignore quality would be a source of embarrassment. In the long run, hopefully, quality in aid interventions would increase and our shared understanding of the world would be better. But would peer reviews, publication and feedback from agencies also alleviate some of the negative effects of research consultancies on local research environments? These were: pulverisation of research ideas, undermining the practice of research as a collective enterprise, and contributing little towards training or capacity-building. In my view, measures such as professional reviews, publication of reports and effective feedback may well have beneficial consequences in all these problem areas. Publication and effective diffusion of consultancy reports may lead researchers interested in applied work to discuss research priorities and their relation to policy formulation. If agencies become more rational in what they commission and expect of studies, tasks set before researchers may become less piecemeal and more conducive to collective efforts. Finally, if considerations of quality become a guiding principle, research consultancies will foster, rather than undermine, the building of local research capacity.

Let us return to the Global Urban Research Initiative. According to its coordinator, among its accomplishments were: (1) that it made the major donors (including the World Bank, some of the agencies, and at least the Ford Foundation) more aware of the in-
teresting work being done on urban subjects by social scientists outside their own structures; (2) that it reinforced certain ideas (such as governance) that were just emerging among the policy community, supporting a more nuanced understanding of what was involved in a number of countries; and (3) that it connected some of the researchers much more effectively into policy networks.3 Hopefully, the findings and insights gained will also inspire foundations, agencies, development banks and other funding bodies to be more constructive in the way they channel resources to social research.

References

3 E-mail communication to Mariken Vaa, 27 April 2003.


