FINNISH ECONOMIC PAPERS Vol. 13, No. 1, Spring 2000

CONTENTS

Finnish Economic Papers – Volume 13 – Number 1 – Spring 2000

BANK MERGERS AND THE FRAGILITY OF LOAN MARKETS*

ERKKI KOSKELA

Department of Economics, Unioninkatu 37, P.O. Box 54, 00014 University of Helsinki, Finland

and

RUNE STENBACKA

Swedish School of Economics, P.O. Box 479, 00101 Helsinki, Finland

We address the question of whether competition makes loan markets more fragile in the sense of increasing the equilibrium bankruptcy risk of firms. This is done using a model of the interaction between the concentration of the banking sector and the investment strategies of imperfectly competitive product market firms. It is shown how a merger between two competing bilateral monopoly banks will typically decrease the interest rate and increase the investment volumes of firms if the investment decisions are strategic complements. Under plausible conditions this implies that a merger will lessen, not aggravate, the fragility of loan markets. (JEL: G21, G33, G34)

1. Introduction

In the 1980s the wave of financial deregulation swept swiftly over many European countries which until then had highly regulated financial markets. The deregulation process was reinforced by one of the main goals of the financial integration within the framework of the European Union, namely to encourage competition in banking.[1] At the same time as the financial integration within the framework of the European Union has encouraged competition in banking, the recent banking crisis in many countries has triggered a wave of substantial bank mergers in the aftermath of financial deregulation. These have been considered as a way for governments to deal with troubled banks and the governments have actively tried

* *This paper is a part of the research program of the Research Unit on Economic Structures and Growth (RUESG) in the department of Economics of the University of Helsinki. The second author would like to acknowledge financial support from Osuuspankkijärjestön Tutkimussäätiö and Bank of Finland. The authors are grateful for helpful comments by Tore Ellingsen, Marco Pagano and Oz Shy.*

[1] *Historically, domestic banks have been protected from competition by stringent regulatory requirements. However, according to the prevailing "home country doctrine" of the European Union, a bank which is registered in any European country is eligible to open branches and to offer financial services in any other member country (see, for example, Mayer and Vives, 1993).*

to encourage mergers of sick banks into healthy ones (see, for example, The Economist, 1995).

In recent times we have seen worldwide an unprecedented high level of merger activity in the banking sector (for updated evidence, see Barfield, 1998). The U.S. banking industry has been consolidating rapidly which has led to increased concentration (see e.g. Boyd and Graham, 1996, for details). As is concluded in the study of The European Central Bank, [ECB] (1999), increased competition can be expected to have a significant impact on the risks incurred by banks. In reflection of a view according to which there would be a tradeoff between competition and financial stability, national authorities in many European countries have approved and, in some cases even encouraged, restructuring of the banking industry whereby the national concentration measures of the banking market have increased in most EU-countries (for evidence, see Tables 3.1–3.3 in ECB, 1999). In Europe mergers among commercial banks have so far taken place predominantly within national markets rather than as cross-border consolidations (see Danthine et. al., 1999, in particular their Tables 3.1 and 3.2, for extensive evidence that the degree of cross-border banking penetration in Europe has remained very modest). The introduction of the EMU is expected to further speed up the consolidation process in the European banking markets.

Conventional wisdom suggests that competition eliminates restrictive practices and reduces margins between borrowing and lending rates thereby improving the performance of the banking industry. The banking industry, however, has several idiosyncratic features, making it hard to evaluate the consequences of increased banking competition as it is not possible to rely directly on general insights from the traditional literature in industrial economics.

A number of recent contributions has analyzed aspects of the relationship between lending market structure and performance of the banking industry. Broecker (1990) and Riordan (1993) have studied the consequences of adverse selection due to the unobserved quality of borrowers. They argue that increased competition may make the adverse selection problems more severe when borrowers that have been rejected at one bank can apply for loans at other banks so that the pool of funded projects show lower average quality as the number of banks increases.[2] Shaffer (1998) has extended the analysis of winner's curse problems in lending and reported empirical evidence about the nature and magnitude of these effects. Contrary to the contributions mentioned above, by identifying the intensity of competition with the degree of product differentiation Villas-Boas and Schmidt-Mohr (1999) have demonstrated how banks facing stronger competition may expose credit applicants to more precise screening under asymmetric information. The central mechanism behind the result of Villas-Boas and Schmidt-Mohr relies on the argument that with stronger competition the banks have to compete more aggressively for the profitable projects. Another approach has investigated some important aspects of the relationship between lenders' market power and the agency cost of debt finance. In a credit-rationing model that take a Schumpeterian R&D perspective Petersen and Rajan (1995) have argued that more intense credit market competition may induce banks to reduce their lending in an initial stage (the asymmetric information period) because they are less able to extract borrower surplus at a subsequent stage (the complete information period) so that lenders in a more competitive market may be forced to initially charge higher interest rates than lenders with more market power.

The present study is focused on the question: How does a particular form of competition in the loan market affect the lending rate charged, the investment decisions of firms in the product market and, in consequence, the default risk of loans? We concentrate on a simple economy with two interdependent firms with access to investment projects in the product market and compare the cases of a banking industry with

[2] *Gehrig (1999) extends this approach within the framework of a model making it possible to explore the relationship between the incentives of banks for costly information acquisition based on ex ante monitoring efforts and the market structure of the lending industry.*

one and two banks, respectively. Firms are assumed to be tied to their banks in such a way that they are effectively prohibited from switching banks so that our analysis focuses on a pair of bilateral monopolies competing with each other. The effects of a banking merger are identified with the consequences for interest rates, investment volumes and bankruptcy risks of a reduction in the number of banks from two to one.[3]

Even though it might be rare to find banking markets with such a high degree of concentration as exhibited in our analysis, the comparison we carry out captures some crucial qualitative implications of bank mergers. Our model seems particularly realistic for countries with banking industries dominated by main-bank relationships and high switching costs. Indeed, there are recent examples which fit our model almost quite literally. The merger between Union Bank of Finland (UBF) and Kansallis-Osake-Pankki (KOP) in early 1995 created a banking giant holding as much as 45% of Finnish deposits and up to 60% of new loans to Finnish small and mid-size corporations. In light of the recent merger cases involving CIBC and Toronto-Dominion as well as Royal Bank of Canada and Bank of Montreal our model should also be relevant from, for example, a Canadian perspective.

Brander and Lewis (1986) initiated research on the linkages between imperfectly competitive product markets and the debt-equity positions of firms. They showed how, in the presence of limited liability, debt induces the firm to an aggressive output strategy. Poitevin (1989) made use of such a strategic relationship to demonstrate how firms belonging to the same imperfectly competitive industry can noncooperatively sustain some degree of collusion by borrowing from the same bank. In Poitevin's model firms borrow in order to finance production of *fixed* size under circumstances where the production decisions are strategic substitutes and where the firms face firm-specific uncertainty. Our analysis differs from Poitevin's in three respects. Firstly, we allow for *endogenously determined loan sizes*. Secondly our analysis captures strategic interaction where the investment decisions are strategic complements. We compare the situation where firms in a duopoly with interrelated projects facing common uncertainty borrow from the same bank with the configuration where firms in a duopoly borrow from two competing banks. Thirdly, we assume that both the firms and the banks face common uncertainty, which enters the profit function of the firms in a multiplicative way (assumption of symmetric information). We postulate a strategic interaction in two stages such that the banks commit themselves to lending rates in the first stage and given these lending rates the firms subsequently decide on their levels of investments in the second stage. Hence we focus on a subgame perfect equilibrium in a two-stage game.

It is shown that under fairly mild conditions a merger between two bilateral monopoly banks would decrease the interest rate and increase the investments of a downstream industry if the investments are strategic complements. An intuitive explanation goes as follows. An increase in the lending rate has two opposite effects on the bank's profits: a negative one, because it induces a reduction of demand, and a positive one, because of the higher return on each unit money used to finance the projects. If the investment decisions of firms are strategic complements, an increase in the lending rate to one of the projects reduces the investment not only for the project in question, but also for the rival's project. Thus, the strategic complementarity strengthens the negative effect associated with an increase in the lending rate. If the banks merge, the resulting monopoly will internalize this effect by charging a lower interest rate than the competing banks. We also provide plausible sufficient conditions under which a bank merger would decrease the fragility of the loan market in the sense of decreasing the equilibrium bankruptcy risks of firms.

The product market interaction between firms investing in risky projects is introduced in section 2. Section 3 explores the implications

[3] *A substantial part of the empirical research on bank mergers has focused on potential economies of scale and scope as a motive for mergers (see e.g. Berger, Hanweck and Humphrey, 1987). Our analysis can be viewed as a complement to this traditional approach.*

of a bank merger on lending rates as well as on total industry investment and delineates the implications for financial fragility by investigating how such a merger will affect the equilibrium bankruptcy risk of firms. The paper ends up with a brief concluding section.

2. Debt-financed risky investments

2.1 Product market interaction

Consider two identical firms, which compete in the product market and both have access to an investment project with an uncertain return. If firm i (i = 1,2) invests x_i while its competitor j invests x_j the project yields $\theta\pi^i (x_i, x_j)$ for firm i provided that the state of nature turns out to be θ. We assume a continuum of possible states of nature θ distributed over the interval $[\theta_L, \theta_H]$ according to a cumulative distribution function $F(\theta)$ with the corresponding density function $f(\theta)$. Thus the firms face common uncertainty, which enters the profit realization in a multiplicative way.[4] The investment technology is presented in the following assumption.[5]

Assumption A: The revenues (net of the stochastic shock θ) of firm i from the project satisfy

(A1) $\pi^i_i(x_i, x_j) > 0$, *(A2)* $\pi^i_{ii}(x_i, x_j) < 0$,
(A3) $\pi^i_i(x_i, x_j) > \left|\pi^i_j(x_i, x_j)\right|$.

Assumptions (A1) and (A2) indicate that the revenue of firm i's project is an increasing and strictly concave function of its own investment. Assumption (A3) captures the idea that "own effects" dominate over "cross-effects".

Since the goal of the present paper is to analyze how competition between bilateral monopoly banks will affect interest rates and investment decisions, and thereby also the risk exposure, of a representative product market industry, we find it justified to assume that the firms have no capital of their own so that the investment projects have to be fully financed with debt from a bank. We will compare two separate banking regime configurations. In the first situation there are two banks, where the two firms in the product market are not able to switch from one bank to another. In the alternative scenario we focus on the case with a bank monopoly offering loans to both the product market duopolists.

Our restriction of competition to that taking place between two competing bilateral monopoly banks in a situation where two firms in the product market are locked into a bank-client relationship should be thought of as a recognition of high switching costs as a central feature of lending markets. Durable relationships between banks and their clients represent a widely observed phenomenon in credit markets.[6] Also the theoretically oriented literature has offered explanations for such a view. For example, Dell'Ariccia (1998) offers a formal model of how informational asymmetries in lending markets can create barriers to entry even in the absence of exogenous fixed costs. Switching costs also play a central role in the important model by Dasgupta and Titman (1996), which deals with the relationship between capital structure and price decisions. Recently, Bolton and Scharfstein (1996) have offered an incomplete contract explanation for concentrating the loans to one common creditor. The firms tend to borrow from one creditor when default risks and asset complementarities are high. Finally, Detragiache et. al (1997) have argued that multiple banking should be less advantageous in countries, where the banking system is more fragile and where the loan recovery process, in the case of a loan default, is more efficient.

---

[4] Dasgupta and Titman (1996) have demonstrated that the consequences of the strategic use of debt might depend on the type and nature of uncertainty facing the product market industry. Since that relationship is outside the focus of our attention, our model exhibits the simplest possible form of common uncertainty.

[5] Partial derivatives are denoted by subscripts.

[6] See e.g. Detragiache, Garella and Guiso (1997) for evidence that U.S. small businesses typically borrow from one bank only. This pattern, however, is not uniform. A striking feature of Italian firms, for instance, is that they ordinarily do business with a variety of different banks.

2.2 Investments with a banking duopoly

In this section we assume that the banks have made their lending rate commitments and we concentrate on the investment decisions of two competing firms maximizing their expected returns. Thus, we analyze the investment equilibrium resulting from the strategic interaction between the firms in the product market, while taking the lending rates as given. Ex ante the firms and the banks face identical aggregate uncertainty, which in our model is separated from the investment technology.

Because the investment is financed with debt, there will be a surplus for the investing firm only when the state of nature is sufficiently good to cover the debt. In a standard debt contract the "breakeven" state of nature η_i, in which firm i is just able to remain solvent, is defined by

$$(1)\quad \eta_i(x_i, x_j, R_i) = \frac{R_i x_i}{\pi^i(x_i, x_j)},$$

where r_i is the interest rate so that $R_i = 1 + r_i$ is the lending factor.

The "breakeven" state of nature depends on the interest rate as well as on the investment level of both the firm itself and of its rival, i.e. $\eta_i = \eta_i(x_i, x_j, R_i)$. The firm remains solvent for states of nature satisfying $\theta \geq \eta_i$, while there is bankruptcy when $\theta < \eta_i$. Consequently, the probability of bankruptcy is given by $F(\eta_i)$. If bank i commits itself to charge a lending factor R_i from firm i, then under the limited liability this firm is protected from negative profits and it will make its investment decision in order to maximize

$$(2)\quad V^i(x_i, x_j) = \int_{\eta_i}^{\theta_H} (\theta \pi^i(x_i, x_j) - R_i x_i)\, dF(\theta).$$

Maximization of (2) (for i=1,2) yields the implicit reaction functions

$$(3)\quad V_i^i(x_i, x_j) = \int_{\eta_i}^{\theta_H} (\theta \pi_i^i(x_i, x_j) - R_i)\, dF(\theta) = 0,$$

which can be rewritten as

$$\int_{\eta_i}^{\theta_H} \theta \pi_i^i(x_i, x_j)\, dF(\theta) = R_i [1 - F(\eta_i)].$$

The left-hand side denotes the marginal revenue increase from an additional unit of investment adjusted to the average among those states of nature in which the firm is solvent. The right-hand side in turn denotes the marginal cost increase in debt from an additional unit of investment for those states of nature, where the firm can fully afford to pay back its debt.

Differentiation of (1) with respect to x_i shows that

$$\frac{\partial \eta_i}{\partial x_i} = \frac{R_i}{\pi^i}(1 - e^{ii}) > 0,$$

because the scale elasticity of investment, $e^{ii}(x_i, x_j)$, satisfies $0 < e^{ii}(x_i, x_j) = \dfrac{x_i \pi_i^i(x_i, x_j)}{\pi^i(x_i, x_j)} < 1$ due to (A2). Analogously, it is easy to show that firm i's "breakeven" state of nature increases with R_i, while the effect of its rival's investment depends on the sign of π_j^i so that we have

$$\frac{\partial \eta_i}{\partial R_i} > 0 \text{ and } sign\left\{\frac{\partial \eta_i}{\partial x_j}\right\} = -sign\,\{\pi_j^i\}$$

Following Clemenz (1986, p. 132) we can rewrite FOC (3) according to

$$(3')\quad \pi_i^i(x_i, x_j)\left[\eta_i + \frac{\int_{\eta_i}^{\theta_H}(1 - F(\theta))d\theta}{1 - F(\eta_i)}\right] - R_i = 0$$

and defining $H(\eta_i) = \dfrac{\int_{\eta_i}^{\theta_H}(1 - F(\theta))d\theta}{1 - F(\eta_i)}$, makes it possible to simplify (3') as

$$(4)\quad R_i(e^{ii}(x_i, x_j) - 1) + \pi_i^i(x_i, x_j) H(\eta_i) = 0.$$

This formulation is convenient as it shows how fairly mild restrictions both on the profit function and on the cumulative distribution function justify the first-order approach.

Lemma 1: FOC (4) is a sufficient condition for optimal investment behavior if *(B1)* $\partial e^{ii}(x_i, x_j)/\partial x_i \leq 0$ and *(B2)* the hazard ratio of F, $f(\eta_i)/[1 - F(\eta_i)]$, is increasing.

For the proof we refer to the Appendix. Assumption (B1) means that the scale elasticity is a non-increasing function of the firm's own investment, which can be regarded as a reasonable assumption. The hazard ratio in Assumption (B2) describes the probability of reaching the breakeven state of nature conditional on not having reached that state earlier.[7]

Corollary 1: Under conditions (A2) and (B2) an increase in the lending rate will decrease the firm's expected marginal return on investment.

Proof: Differentiating (4) with respect to the interest factor gives

$$(5)\quad V^i_{iR_i} = (e^{ii}(x_i, x_j) - 1) + \pi^i_i(x_i, x_j) H'(\eta_i) \frac{\partial \eta_i}{\partial R_i} < 0.$$

The sign of (5) follows from the concavity of the profit function ($e^{ii} < 1$), from (B2) and from the fact that $\partial\eta/\partial R_i > 0$. *QED*

Provided that conditions (B1) and (B2) hold, the Nash equilibrium in investments is characterized by the system of equations

$$(6)\quad V^i_i(x^*_i, x^*_j) = 0$$

$$(7)\quad V^j_j(x^*_j, x^*_i) = 0.$$

(6) and (7) are the equilibrium conditions for given levels of the interest factors R_i and R_j. In what follows our analysis of the industry equilibrium is restricted to situations which satisfy the standard conditions for stability so that there is a unique investment equilibrium. Technically these conditions are $\Delta_V = V^i_{ii} V^j_{jj} - V^i_{ij} V^j_{ji} > 0$, where $V^i_{ii} < 0$ (i=1,2) due to the second-order conditions.

By totally differentiating the system of equations defined by (6) and (7) with respect to R_i we find that

$$(8)\quad \frac{\partial x^*_i}{\partial R_i} = -\frac{V^j_{jj} V^i_{iR_i}}{\Delta_V}$$

and

$$(9)\quad \frac{\partial x^*_j}{\partial R_i} = -\frac{V^j_{ji}}{V^j_{jj}} \frac{\partial x^*_i}{\partial R_i}.$$

For the analysis of (8) and (9) we differentiate the first-order condition (4) with respect to x_j to find that

$$(10)\quad V^i_{ij} = R_i e^{ii}_j + \pi^i_{ij} H(\eta_i) - \pi^i_i H'(\eta_i) \frac{R_i x_i \pi^i_j}{(\pi^i)^2}.$$

We employ the following assumption.

Assumption C: The scale elasticity of firm i is a non-decreasing function of the investment level of the rival firm j so that

$$(C1)\quad e^{ii}_j = \frac{\partial e^{ii}(x_i, x_j)}{\partial x_j} = \frac{x_i(\pi^i \pi^i_{ij} - \pi^i_i \pi^i_j)}{(\pi^i)^2} \geq 0.$$

Condition (C1) means that a firm's relative marginal return on investment does not decrease as the rival's investment increases. For example, a combination of conditions $\pi^i_j \leq 0$ and $\pi^i_{ij} \geq 0$ would be sufficient for (C1) to hold. Also a Cobb-Douglas function of the form $\pi^i(x_i, x_j) = x_i^\beta x_j^{1-\beta}$, where $0 \leq \beta \leq 1$, satisfies condition (C1).

Lemma 2: If the investments are strategic complements ($\pi^i_{ij} > 0$) and if conditions (B2) and (C1) hold, firm i's expected marginal return on investment is an increasing function of its rival's investment ($V^i_{ij} > 0$).

Proof: Follows directly from combining (10) and Lemma 1. *QED*

Consequently, conditions (B2) and (C1) are sufficient to make sure that we can extend the property of the investments being strategic complements in the ordinary sense to hold also in situations where these investments are financed by debt under limited liability.

One can distinguish between technological complementarities and complementarities induced by to aggregate demand externalities. The standard duopoly R&D race discussed e.g. in Harris and Vickers (1987) is a good example of a model where the investment decisions are strategic complements due to technological externalities[8] (see also Bagwell and Staiger, 1994,

[7] *Assumption (B2) is commonly used, for example, in the principal-agent literature as the monotone likelihood-ratio condition (see Rogerson, 1985), and as Jorgensen, McCall and Radner (1967) show, it is satisfied for a wide range of distribution functions.*

[8] *In a paper slightly related to ours, Clemenz (1991) has focused on some aspects which are of relevance for projects with interrelated returns. His article has a particular emphasis on R&D projects.*

for other characterizations). A potentially even more convincing justification for the assumption of strategic complementarities is based on aggregate demand externalities, whereby an expansion of firm i's investments would benefit firm j. Such aggregate demand externalities are particularly relevant under circumstances with imperfectly competitive firms which are large relative to the size of the economy (see Matsuyama, 1995). One can consider our model to capture the aggregate demand externalities in a reduced form. In a recent study, controlling for industry effects and time effects, Stenbacka and Tombak (1999) offer empirical evidence across a large number of industries which supports the notion that the investment decisions of large competing firms are strategic complements.

We are now able to evaluate (8) and (9). Based on Lemma 2 and the assumption of stability it must hold that

$$(11)\quad \frac{\partial x_i^*}{\partial R_i} < \frac{\partial x_j^*}{\partial R_i} < 0.$$

Thus, an increase in the lending rate levied on firm i will reduce the investment of both firm i and its rival because of strategic complementarity between their investments. But, as a result of the stability condition, the reduction in firm i's own investment will be larger than that of its rival.

2.3 *Investments with a banking monopoly*

Since we intend to compare the investment sensitivity to lending rate changes between a banking monopoly and duopoly we now turn to analyze a bank monopoly. A banking monopoly finds it optimal to charge the identical duopolists in the product market a common lending factor R. Keeping the notation otherwise unchanged, firm i will then decide on its investment level in order to maximize

$$(12)\quad V^i(x_i, x_j) = \int_{\eta_i}^{\theta_H} (\theta \pi^i(x_i, x_j) - Rx_i)\, dF(\theta).$$

Following the approach of the previous section we can now characterize the impact of the lending rate on the equilibrium investment decisions. Total differentiation of the first-order conditions for the investments reveals that

$$(13)\quad \frac{\partial x_i^*}{\partial R} = -\frac{V_{iR}^i (V_{ii}^i - V_{ji}^j)}{\Delta_V} < 0,$$

where we have made use of (5) together with Lemmas 1 and 2. The terms in the numerator of (13) refer to second order partial derivatives of the objective function (12) and their properties are analogous to those in the bank duopoly case. The difference factor $V_{ii}^i - V_{ji}^j$ in (13) captures that a monopoly bank internalizes the strategic externalities between the two product market duopolists it finances and we know that under the assumptions made this difference is negative. The denominator Δ_V refers to the sufficient second-order and stability conditions for the maximization of (12) (i = 1,2) and it satisfies that $\Delta_V > 0$.

Under the assumptions prevailing, comparison of (8), (9) and (13) makes it possible to formulate the relationship

$$(14)\quad \frac{\partial x_i^*}{\partial R} < \frac{\partial x_i^*}{\partial R_i} < \frac{\partial x_j^*}{\partial R_i} < 0,$$

if the investment decisions are strategic complements. Thus, an increase in the monopoly lending rate will reduce industry investment to a greater extent than a corresponding increase in the lending rate charged by the banks in a duopolistic industry. This can be explained as follows. The lending rate charged by a monopoly bank affects investment behavior of the firms directly and symmetrically. But, when a bank in a duopolistic industry increases its lending rate it will also affect the investment behavior of its customer's rival. An increase in the lending rate confronting firm i, however, generates a smaller investment change for the rival firm j than for firm i.

3. *Interest rate decisions*

Having analyzed the investment decisions of firms and their reactions to changes in the lending rates, we now turn to consider the first stage of the game, i.e. the lending rate decisions of banks. We start by looking at a banking duopoly and then move on to the case with a bank monopoly. Finally, we examine the implications of a bank merger on the fragility of the banking industry by investigating how such a merger will affect the equilibrium bankruptcy risk of firms.

3.1 Interest rate decisions in duopoly and monopoly banking

The banks make the lending rate commitments taking into account how the interest rates will affect the investments of the firms in the product market. In the previous section we delineated the investment equilibrium for given lending rates which results from the strategic interaction between the firms in the product market. Given this equilibrium, bank i commits itself to lend to firm i at an interest factor R_i which maximizes the expected value of the debt contract from bank i

$$(15)\quad \Gamma^i(R_i, R_j) = \pi^i(x_i^*, x_j^*)\int_{\theta_L}^{\eta_i}\theta dF(\theta) + x_i^*[(1-F(\eta_i))R_i - R_0],$$

where $R_0 = 1 + r_0$ denotes the opportunity cost factor of granting loans, which is assumed to be constant. The objective function (15) reflects the assumption that the bank knows the nature of the investment technology available. The first term on the right hand side of (15) describes the bank's profit in those states of nature where firm i faces bankruptcy. The second term expresses the bank's profits net of the opportunity cost of granting loans in those states of nature where the firm remains solvent.

Maximization of (15) (for i = 1,2) yields the implicit reaction functions for duopoly banks

$$(16)\quad -\left[\frac{\partial\pi^i(x_i^*, x_j^*)}{\partial x_i}\frac{\partial x_i^*}{\partial R_i} + \frac{\partial\pi^i(x_i^*, x_j^*)}{\partial x_j}\frac{\partial x_j^*}{\partial R_i}\right]\int_{\theta_L}^{\eta_i}\theta dF(\theta) = x_i^*[1-F(\eta_i)] + \frac{\partial x_i^*}{\partial R_i}[(1-F(\eta_i))R_i - R_0],$$

where the first-order condition has been simplified based on (1). The first term in squared brackets on the left-hand side of (16) represents a direct negative effect of an increase in the interest rate factor, because a lending rate increase reduces the gross return of investment and thereby the profit of the lender in those states of nature where the firm defaults. The second term in squared brackets on the left-hand side of (16) captures the indirect effect due to the strategic investment interaction between the projects. By combining Assumption (A3) with (11) we can conclude that this indirect effect cannot dominate the direct effect. Hence, the left-hand side of (16) is positive. The first term on the right-hand side of (16) describes the direct interest rate effect in those states of nature where the firm remains solvent. The second term on the right-hand side of (16) expresses the indirect effect induced by changes in investments in those states of nature where the firm remains solvent. The intersection between the reaction functions defined by (16) will constitute the lending rate equilibrium between two competing bilateral monopoly banks.

In order to distinguish the interest rate equilibrium in a banking duopoly from the optimal lending rate of a monopoly bank we have to study also the lending rate determination for the monopoly bank. A monopoly bank will choose R so as to maximize

$$\Psi(R) = \pi^i(x_i^*, x_j^*)\int_{\theta_L}^{\eta_i}\theta dF(\theta) + x_i^*[(1-F(\eta_i))R - R_0].$$

Since an increase in the interest rate will have a symmetric effect on the investments of the competing firms, it directly follows that it is optimal for such a monopoly bank to charge the same lending rate from the firms. Therefore the first-order condition can be written as

$$(17)\quad -\left[\frac{\partial\pi^i(x_i^*, x_j^*)}{\partial x_i}\frac{\partial x_i^*}{\partial R} + \frac{\partial\pi^i(x_i^*, x_j^*)}{\partial x_j}\frac{\partial x_j^*}{\partial R}\right]\int_{\theta_L}^{\eta_i}\theta dF(\theta) = x_i^*[1-F(\eta_i)] + \frac{\partial x_i^*}{\partial R}[(1-F(\eta_i))R - R_0],$$

The interpretation of (17) is analogous to that of (16) except for the fact that (17) is an ordinary first-order condition while (16) defines the reaction functions. Conditions (16) and (17) allow us to compare the optimal lending rate of a bank monopoly with the interest rate equilibrium prevailing in a banking duopoly.[9] Now one can establish

Proposition 1: When the investment decisions are strategic complements, a merger of duopoly banks into a monopoly bank will decrease the lending rate.

Proof: For the proof, see the appendix[10].

An explanation goes as follows. A rise in the lending rate has two opposing effects on the bank's profits: a negative one, because it induces a reduction in the investment volumes, and a positive one, because of the higher return on each unit money used to finance the projects. If the investment decisions of firms are strategic complements, an increase in the lending rate to one of the projects reduces the investment not only for the project in question, but also for the rival's project. Thus, the strategic complementarity strengthens the negative effect associated with an increase in the lending rate. If the banks merge, the resulting monopoly will internalize this effect by charging a lower interest rate than the competing banks.

Proposition 1 has an immediate corollary as one can see from equation (14).

Corollary 2: A bank merger will imply an expansion of the investment programs.

Proposition 1 and Corollary 2 show potential gains for interdependent firms from concentrating their borrowing activities to a single monopoly bank. We have identified competition with the number of banks. This could be justified by arguing that more competition is plausibly related to a higher degree of fragmentation of the banking industry and hence it could conceivably happen that interdependent firms are financed by different lenders with a higher probability when the degree of competition is high.[11]

Throughout the analysis we have focused on the consequences of a merger whereby the banking industry is transformed from a duopoly to a monopoly. However, the same arguments would hold in response to any change in the banking structure from n to n–1 banks ($n \geq 2$). Clearly, the impact of such a change is most dramatic, and can therefore be demonstrated in the strongest and simplest possible way, under the scenario which is exhibited in our analysis. Mechanisms whereby informational asymmetries in lending markets can create a barrier to entry for new lending institutions even in the absence of exogenous fixed costs offer an additional justification for why we formally focus on lending market structures which are very concentrated. Dell'Ariccia (1998) characterize one such mechanism, where creditworthy borrowers, unable to signal their quality to competing lenders, are locked into bank-credit relationships of the type considered in our model.

Proposition 1 is restricted to hold only when the investment decisions are strategic complements. In order to see that it cannot be generalized to hold when the investment decisions are strategic substitutes we consider the following simple example. Assume that the firms make production investments within the framework of a Cournot model with linear inverse demand $p = 1 - x_i - x_j$, where x_i denotes the production investment of firm i. Suppose further for simplicity that there is no uncertainty. In such a case the profit of firm i is given by $\pi^i(x_i, x_j) = x_i\,[1 - x_i - x_j - R_i]$. For given lending factors, R_i and R_j, the investment equilibrium (the Cournot equilibrium) can then be calculated to be $x_i^* = \frac{1}{3}[1 - 2R_i + R_j]$. Let us further normalize so

[9] *It is assumed that the complicated sufficient second-order and stability conditions are satisfied in a way analogous to the conditions presented in section 2.2.*

[10] *Our proof for this is different and applies under much more general conditions than the proof carried out by Poitevin.*

[11] *Berger, Kashyap and Scalise (1995) have studied what has happened to the size distribution of bank loans in U.S. after bank mergers due to the relaxation of interstate banking. According to their results there has been a loss of market share by small banks in favour of large banks. This finding lies in conformity with our analysis.*

that the opportunity cost of loans is zero, i.e. $R_0=0$. In the duopoly case bank i then determines the lending rate in order to maximize $\Gamma^i(R_i, R_j)=\frac{1}{3}[1-2R_i+R_j]R_i$. Straightforward calculations show that the lending rate equilibrium is then $R_i^*=\frac{1}{3}$. In case a bank merger takes place the resulting lending market monopoly would find it optimal to raise the interest factor to $R^m=\frac{1}{2}$. Consequently, when the investment decisions are strategic substitutes the merged bank would internalize the strategic interaction between the interdependent loan projects by increasing the lending rates.

The feature with a common lender deciding on a lower interest rate is present also in Poitevin's (1989) analysis of the impact of lending rates on production decisions which are strategic substitutes. He explains the mechanism by referring to a lower interest rate as a way for a bank monopoly to limit firm's incentives to choose risky output. In our model with endogenously determined loan volume the interpretation of the interaction between financial structure and imperfect product market competition is different, because Poitevin's analysis is restricted to the case where firms have to debt-finance a fixed expenditure (normalized to one) in order to produce.

Next we will turn to investigate the relationship between the lending market structure and the fragility.

3.2 Bank merger and equilibrium bankruptcy risk

In the case of the Scandinavian banking crisis in the early 90's it has often been pointed out that the crisis was preceded by deregulation of the banking market. For that reason many observers have used the Scandinavian experience as evidence of how increased competition in the bank loan market will generate higher instability in the financial sector.[12] When evaluating the impact of bank mergers on financial stability the crucial issue is how such a consolidation of the banking industry will affect the failure rates. In this section these issues are investigated from a restricted point of view where we make no attempt to explicitly model the mechanisms whereby borrowers' project failures generate states of bank insolvency. In this respect we simply subscribe to the view expressed by Bhattacharya, Boot and Thakor (1998) insofar as "business activity measures, such as small business failure rates, are very useful in predicting bank runs" (p. 752). As long as the borrowers are homogeneous the identification of credit market fragility with failure rates of borrowing firms can hardly be seen as very restrictive. Of course, the transmission mechanism from business failure rates to bank solvency is more complex with heterogeneous borrowers. Our model, however, is too complicated for a detailed analysis of this interesting mechanism.

In the previous subsection we established how under certain conditions a bank merger would decrease the interest rate and increase the level of investments. This suggests that the effect of a bank merger on the fragility of loan markets, in the sense of equilibrium bankruptcy risks of borrowers, is not a priori clear. While a fall in the interest rate decreases the default risk, the resulting increase in investment does the reverse.

To study this issue more closely we consider the equation

$$(18)\quad \eta_i^* \pi^i(x_i^*, x_j^*) - R_i x_i^* = 0,$$

where η_i^* denotes the "breakeven" state of nature generated by equilibrium investments and interest rates. Differentiating η_i^* as defined by (18), we prove in the Appendix that

$$(19)\quad \frac{\partial \eta_i^*}{\partial R_i} = \frac{x_i^*}{\pi^i}\,[1-\varepsilon^*(1-(e^{ii}+\lambda e^{ij}))],$$

in which $\varepsilon^* = -\dfrac{\partial x_i^*/x_i^*}{\partial R_i/R_i} > 0$ denotes the interest rate elasticity of investment, while $e^{ii}+\lambda e^{ij}$ is the adjusted scale elasticity with respect to industry investments. The adjusted scale elastici-

[12] *According to Honkapohja and Koskela (1999) the credit losses of the Finnish banks amounted on average to 4.1% of the amount of lending during the period 1991–93, while the corresponding figure for Sweden was 4.3%.*

ty with respect to industry investments is defined by $e^{ij}=\frac{\pi_j^i x_j}{\pi^i}$ together with $\lambda=\frac{\partial x_j^*/\partial R_i}{\partial x_i^*/\partial R_i}$ $=-\frac{V_{ji}^j}{V_{jj}^j}$ so that $0 < \lambda < 1$ based on Lemma 2. The adjustment factor λ accounts for the fact that an increase in the lending rate for firm i will change its investment to a larger extent than that of its rival. We can thus conclude that

(20a) $\frac{\partial \eta_i^*}{\partial R_i} > 0 \text{ for all } \varepsilon^* \text{ when } e^{ii}+\lambda e^{ij} \geq 1$

and

(20b) $\frac{\partial \eta_i^*}{\partial R_i} > 0 \text{ iff } \varepsilon^* < \frac{1}{1-(e^{ii}+\lambda e^{ij})}$

$\text{when } e^{ii}+\lambda e^{ij} < 1.$

Consequently, (20a) and (20b) delineate the conditions under which the equilibrium "breakeven" state of nature η^* depends positively on the lending rate. Under such conditions a lending rate increase will raise the equilibrium probability of default because $F'(\eta) > 0$.

Equation (19) decomposes the impact of the lending rate on bankruptcy into three different effects. Firstly, if the investment volume does not respond to a lending rate change, as in Poitevin (1989), only the first term would be relevant and a merger would automatically stabilize the loan market, since then the lending rates would directly determine the equilibrium probability of firm bankruptcy. However, a change in the interest rate facing firm i will typically also generate effects on the investment volumes of both firms i and j. In the absence of strategic interaction between the funded projects we would have an investment response working in an opposite direction to the 'direct effect', but as long as $\varepsilon^* < [1-e^{ii}]^{-1}$ the 'direct effect' would dominate. The presence of strategic interaction between the firms will further modify the investment response to interest rate changes. The investment response to a merger will be strengthened (weakened) when $\lambda e^{ij} < 0$ ($\lambda e^{ij} > 0$).

As a merger between two bilateral monopoly banks will decrease the equilibrium interest rate according to Proposition 1 we can formulate

Proposition 2: A merger between two bilateral monopoly banks will lessen the fragility of loan markets by reducing the bankruptcy risk of firms (a) always when the adjusted scale elasticity with respect to industry investments weakly exceeds one, (b) if and only if the interest rate elasticity of investment is "sufficiently small"**,** when the adjusted scale elasticity with respect to industry investments is below one.

An intuitive interpretation is as follows. When the response of the industry investment is greater than or equal to one, i.e. when $e^{ii} + \lambda e^{ij} \geq 1$, the investment effect will add on to (or does not counteraffect) the 'direct' interest rate effect in stabilizing the loan market, because an interest rate increase will add to the bank's revenues (or leave them unchanged) via the generated investment increase. Consequently, a merger between duopoly banks will reduce the fragility of the credit market.

On the other hand, when $e^{ii} + \lambda e^{ij} < 1$ the generated investment increase will decrease the bank's revenues via the generated investment growth. The literature offers some guidance regarding the quantitative size of the response of firms' investment decisions to changes in interest rates. With a production function with capital and labor of Cobb-Douglas type, the interest rate elasticity would be equal to one. This is regarded as an upper bound for the interest rate elasticity (see Abel, 1990, p. 762–63, for more details). In fact, the huge empirical literature on investment functions suggests that the elasticity is well below one (see Chirinko, 1993, for a recent survey). Thus as an empirical matter we can safely expect the condition in Proposition 2 (b) to hold even when $e^{ii} + \lambda e^{ij} < 1$.

The present section has demonstrated that a bank merger does not increase the fragility of the credit market despite the fact that such a merger will generate a reduction in the lending rate under the circumstances considered. In this respect the present model reinforces the central argument in Koskela and Stenbacka (2000), which exhibits a completely different mechanism for the absence of a tradeoff between low lending rates and high failure rates of projects funded with debt. In Koskela and Stenbacka

(2000) uncertainty is generated conditional on the investments, making thus uncertainty an endogenous feature of that model.

Propositions 1 and 2 can be used to explore some important welfare implications of a bank merger within the context of our model with investments being strategic complements. As a benchmark for a welfare evaluation we formulate the following social objective function:

$$(20)\quad W(x_i(R),x_j(R)) = \big(\pi^i(x_i(R),x_j(R)+\pi^j(x_j(R),x_i(R))\big)\int_0^{\theta_H}\theta dF(\theta) - R\big(x_i+x_j\big)-\big(F(\eta^i(x_i,x_j,R)) + F(\eta^j(x_j,x_i,R))\big)K,$$

where K denotes a fixed social cost of bankruptcy and where we consider a symmetric equilibrium ($R = R_i = R_j$). The social objective function (20) captures the idea that society is indifferent regarding the distribution of resources between lenders and borrowers and also incorporates social costs of bankruptcy in the simplest possible way. In order to find out the welfare consequences of a bank merger which according to Propositions 1 and 2 both decreases the lending rate and the failure rate of borrowers, we totally differentiate (20) with respect to R. Proceeding with such an approach we find that

$$(21)\quad \frac{dW(x_i, x_j)}{dR} = 2\left[E(\theta)(\pi_i^i+\pi_i^j)\frac{dx_i}{dR} - x_i(1-\varepsilon) - \frac{dF(\eta^i)}{dR}\mathrm{K}\right] < 0,$$

where $E(\theta) = \int_0^{\theta_H}\theta dF(\theta) > 0$, $\pi_i^i + \pi_i^j < 0$ by (A3), $\varepsilon \leq 1$, as previously, denotes the interest rate elasticity of investment and where it holds that $dF(\eta^i)/dR > 0$ by Proposition 2. Consequently, the feature that a bank merger reduces failure rates of borrowers adds on to the welfare gains, which are created by expanded investments due to a lower interest rate.

4. Concluding discussion

We have modelled the interaction between the concentration of the banking sector and the investment strategies of imperfectly competitive firms in the product market to address the question of whether competition makes loan markets more fragile by increasing the equilibrium bankruptcy risk of firms. It has been shown how under fairly mild conditions a merger between two bilateral monopoly banks would decrease the interest rate and increase the investment volume of imperfectly competitive firms if the investment decisions are strategic complements. Under quite plausible conditions our model implies that a merger would lessen the fragility of loan markets. With fixed social costs of bankcruptcy reduced failure rates add on to the welfare gains which are created by expanded investments due to a lower lending rate in equilibrium.

This paper has delineated the advantages for interdependent firms of borrowing from the same monopoly bank rather than from duopolists in the banking market. These advantages could be compared with those of a common agent as identified by Bernheim and Whinston (1985). However, as in Poitevin (1989), our results are different from theirs, because a common monopoly bank will not be able to dictate the investment decisions of the borrowing firms and this weakens the collusive power of a common lender in the banking market.

Throughout our analysis we have restricted our attention to lending markets by assuming that the bank’s opportunity cost of granting loans is constant. Clearly, as financial intermediaries banks compete simultaneously both in the output (loan) and input (deposit) markets. Within such a general framework the opportunity cost of funds is itself endogenous (as elaborated in the contribution by, for example, Yanelle, 1997). In order to more extensively evaluate the consequences of bank mergers it is also important to explore their implications with respect to the performance of the deposit market. The present model has delineated circumstances under which a merger between two duopoly banks will decrease the lending rate without increasing the fragility of loan markets in the sense of increasing the bankruptcy risks of firms in equilibrium. Insofar as the circumstances outlined imply that the bankruptcy risks of a merged bank are reduced, such a merger is likely to generate a comparative advantage also with respect to the opportunity cost of loan funding. Namely, it is plausible that lower bankruptcy risks make the merged bank able to attract risk averse depositors at lower deposit rates. In this respect the main results of the present analysis seem to be robust relative to generalizations incorporating deposit markets.

Our model has formally been restricted to the case of common uncertainty meaning that the risks facing the investments are strictly correlated across the different projects. However, when evaluating the implications of assuming that the risks of the projects are strictly correlated it should be emphasized that it is an assumption which actually is in disfavour of a bank merger because in the presence of such an assumption a merger does not achieve risk diversification. Consequently, in this respect our assumptions rather underestimate than overestimate the potential gains from bank mergers.

The present analysis has abstracted from investigating how introduction of competition might affect the diversification incentives of banks. In light of the recent study by Shy and Stenbacka (1999) we have reasons to conjecture that our results are robust to generalizations making the diversification decisions of the banks an endogenous variable. Shy and Stenbacka demonstrate how, in the context of a differentiated banking industry, the banks' incentives for diversification are invariant to a change in the banking market structure from duopoly to monopoly.

A central conflict of interests between shareholders and bondholders is a typical feature of debt contracts. The shareholders place emphasis only on states of nature that are solvent, while in bankrupt states the shareholders' losses are truncated at zero due to limited liability. Debtholders, on the other hand, place emphasis only on bankrupt states, which would distort them to favor investment strategies which are too conservative relative to the first-best level. These observations led Stiglitz (1985) to suggest debtholder representation in the boards of borrowing firms as a mechanism of implementing first-best investment levels. Later on Brander and Poitevin (1992) have presented a much more detailed model of the agency costs of debt, in which they showed how the terms of the compensation contract offered to outside management by shareholders can reduce these agency costs substantially. Our results suggest that the agency costs of debt are dependent on the market structure in the banking industry, which is an important subject to further research.

References

Abel, A.B. (1990). "Consumption and Investment." In *Handbook of Monetary Economics*, Vol. 2. Eds. B.M. Friedman and F.H. Hahn. Amsterdam: North-Holland.

Bagwell, K., and R. Staiger (1994). "The Sensitivity of Strategic and Corrective R&D Policy in Oligopolistic Industries." *Journal of International Economics* 36, 133–150.

Barfield, R. (1998). "Creating Value Through Mergers." *The Banker*, July 1998.

Berger, A.N., A.K. Kashyap, and J.M. Scalise (1995). "The Transformation of the U.S. Banking Industry: What a Long, Strange Trip It's Been." *The Brookings Papers on Economic Activity* 2, 55–218.

Berger, A.N., G..A. Hanweck, and D.B. Humphrey (1987). "Competitive Viability in Banking: Scale, Scope and Product Mix Economies." *Journal of Monetary Economics* 20, 501–520.

Bernheim, B.D., and M.D. Whinston (1985). "Common Market Agency as a Device for Facilitating Collusion." *Rand Journal of Economics* 16, 269–281.

Bhattacharya, S., A. Boot, and A. Thakor (1998). "The Economics of Bank Regulation." *Journal of Money, Credit and Banking* 30, 745–770.

Bolton, P., and D.S. Scharfstein (1996). "Optimal Debt Structure and the Number of Creditors." *Journal of Political Economy* 104, 1–25.

Boyd, J., and S. Graham S. (1996). "Consolidation in U.S. Banking: Implications for Efficiency and Risk." Research Department Working Paper 572, Federal Reserve Bank of Minneapolis.

Brander, J.A., and T.R. Lewis (1986). "Oligopoly and Financial Structure: The Limited Liability Effect." *American Economic Review* 76, 956–970.

Brander, J.A., and M. Poitevin (1992). "Managerial Compensation and the Agency Cost of Debt Finance." *Managerial and Decision Economics* 13, 55–64.

Broecker, T. (1990). "Credit-Worthiness Tests and Interbank Competition." *Econometrica* 58, 429–452.

Chirinko, R. S. (1993). "Business Fixed Investment Spending: A Critical Survey of Modelling Strategies, Empirical Results, and Policy Implications." *Journal of Economic Literature* 31, 1875–1911.

Clemenz, G. (1986). *Credit Markets with Asymmetric Information*. Berlin: Springer Verlag.

– **(1991).** "Loans for Projects With Interdependent Returns." *Journal of Institutional and Theoretical Economics* 147, 337–353.

Danthine, J.-P., F. Giavazzi, X. Vives, and E.-L. von Thadden (1999). *The Future of European Banking.* Monitoring European Integration Series 9, Centre for Economic Policy Research, London.

Dasgupta, S., and S. Titman (1996). "Pricing Strategy and Financial Policy" NBER Working Paper 5498.

Dell'Ariccia, G. (1998). "Asymmetric Information and the Structure of the Banking Industry" IMF – Research Department Working Paper 98/92. Forthcoming in Journal of Finance.

Detragiache, E., P.G. Garella, and L. Guiso (1997). "Multiple versus Single Banking Relationships" CEPR Discussion Paper No. 1649.

The Economist (1995). Please, Governor, Can You Spare a Billion?, March 25, 99–100.

European Central Bank [ECB] (1999). *Possible Effects of EMU on the EU Banking Systems in the Medium to Long Term.*

Gehrig, T. (1999). "Screening, Cross-Border Banking, and the Allocation of Credit." *Research in Economics*, 52, 387–407.

Harris, C., and J. Vickers (1987). "Racing with Uncertainty." *Review of Economic Studies*, 54, 1–21.

Honkapohja, S., and E. Koskela (1999). "Finland's Depression: A Tale of Bad Luck and Bad Policies." *Economic Policy,* October, 29, 399–436.

Jorgensen, D.W., J.J. McCall, and R. Radner (1967). *Optimal Replacement Policy.* Chicago: Rand McNelly Publ. Co.

Koskela, E., and R. Stenbacka (2000). "Is There a Tradeoff between Bank Competition and Financial Fragility?" *Journal of Banking and Finance*, forthcoming.

Matsuyama, K. (1995). "Complementarities and Cumulative Processes in Models of Monopolistical Competition." *Journal of Economic Literature* 33, 701–729.

Mayer, C., and X. Vives (eds.) (1993). *Capital Markets and Financial Intermediation.* Cambridge: Cambridge University Press.

Petersen, M., and R. Rajan (1995). "The Effect of Credit Market Competition on Lending Relationships." *Quarterly Journal of Economics* 110, 407–443.

Poitevin, M. (1989). "Collusion and the Banking Structure of a Duopoly." *Canadian Journal of Economics* 89, 263–277.

Riordan, M. (1993). "Competition and Bank Performance: A Theoretical Perspective." In *Capital Markets and Financial Intermediation.* Eds. C. Meyer and X. Vives. Cambridge: Cambridge University Press, 328–343.

Rogerson, W. (1985). "The First-Order Approach to Principal-Agent Problems." *Econometrica* 53, 1357–1367.

Shaffer, S. (1998). "The Winner's Curse in Banking." *Journal of Financial Intermediation* 7, 359–392.

Shy, O., and R. Stenbacka (1999). "Market Structure and Risk Taking in the Banking Industry." Working paper, University of Haifa.

Stenbacka, R., and M. Tombak (1999). "Strategic Investments: Theory and Empirical Evidence." Mimeo, Helsinki School of Economics.

Stiglitz, J.E. (1985). "Credit Markets and Control of Capital." *Journal of Money, Credit and Banking* 17, 133–152.

Villas-Boas, M., and U. Schmidt-Mohr (1999). "Oligopoly with Asymmetric Information: Differentiation in Credit Markets." *Rand Journal of Economics* 30, 375–396.

Yanelle, M.O. (1997). "Banking Competition and Market Efficiency." *Review of Economic Studies* 64, 215–239.

Appendix

Proof of Lemma 1: Differentiation of (4) yields

$$V^i_{ii}(x_i, x_j) = R\frac{\partial e^{ii}}{\partial x_i} + \pi_{ii}(x_i, x_j)H(\eta_i) + \pi^i_i(x_i, x_j)H'(\eta_i)\frac{\partial \eta_i}{\partial x_i}.$$

As it can be shown, an increasing hazard ratio of F is sufficient for the property $H'(\eta_i) < 0$. Since $\pi^i_{ii} < 0$, $\partial\eta_i/\partial x_i > 0$ and $\pi^i_i > 0$ it follows that (B1) and (B2) are sufficient to justify the first-order approach. *QED*

Proof of Proposition 1: Let us define the function $g^i(R_i, R_j)$ according to

$$g^i_i(R_i, R_j) = \frac{\partial x^*_i}{\partial R_i}[(1-F(\eta_i))R_i - R_0] + x^*_i[1-F(\eta_i)].$$

Consider the system of equations

$$(21) \quad g^i_i(R_i, R_j) + \alpha\left[\frac{\partial \pi^i(x^*_i, x^*_j)}{\partial x_i} + \frac{\partial \pi^i(x^*_i, x^*_j)}{\partial x_j}\right]\int_{\theta_L}^{\eta_i}\theta dF(\theta) = 0.$$

In a symmetric equilibrium we get that $R_i = R_j$, and, consequently, (16) is also an equilibrium condition. However, (16) can also formally be interpreted as the optimality condition of a merged bank charging the same interest rate to both borrowers. From the relationships (14) we know that a bank merger corresponds to a decrease in α if the investment decisions are strategic complements and if Assumption (A3) holds. In order to find out the impact of a bank merger on the interest rate we totally differentiate the system of equations (21) with respect to α. Leaving out the arguments from the function g^i, total differentiation yields

$$g^j_{jj}\frac{\partial R_j}{\partial \alpha} + g^j_{ji}\frac{\partial R_i}{\partial \alpha} + \left[\frac{\partial \pi^j(x^*_j, x^*_i)}{\partial x_j} + \frac{\partial \pi^j(x^*_j, x^*_i)}{\partial x_i}\right]\int_{\theta_L}^{\eta_i}\theta dF(\theta) = 0$$

and, similarly with respect to firm j,

$$g^j_{jj}\frac{\partial R_j}{\partial \alpha} + g^j_{ji}\frac{\partial R_i}{\partial \alpha} + \left[\frac{\partial \pi^j(x^*_j, x^*_i)}{\partial x_j} + \frac{\partial \pi^j(x^*_j, x^*_i)}{\partial x_i}\right]\int_{\theta_L}^{\eta_i}\theta dF(\theta) = 0.$$

Solution of this system of equations shows that

$$\frac{\partial R_i}{\partial \alpha} = -\frac{g^j_{jj} + g^i_{ij}}{g^i_{ii}g^j_{jj} - g^i_{ij}g^j_{ji}}\left[\frac{\partial \pi_i(x^*_i, x^*_j)}{\partial x_i} + \frac{\partial \pi^i(x^*_i, x^*_j)}{\partial x_j}\right]\int_{\theta_L}^{\eta_i}\theta dF(\theta) > 0.$$

This conclusion is based on a combination of ordinary sufficient second order conditions ($g^i_{ii} < 0$), stability conditions ($g^i_{ii}g^j_{jj} - g^i_{ij}g^j_{ij} > 0$) and on the assumption that "own effects" dominate over "cross-effects" (A3) (implying that $g^i_{ii} - g^i_{ij} < 0$). Consequently, we have proved that $\partial R_i/\partial\alpha > 0$, from which the conclusion of the proposition follows. *QED*

Proof of (19): Differentiating (18) we find that

$$(22) \qquad \frac{\partial \eta_i^*}{\partial R_i} = \frac{x_i^* + R_i \dfrac{\partial x_i^*}{\partial R_i} - R_i x_i^* \left(\dfrac{\pi_i^i}{\pi^i} \dfrac{\partial x_i^*}{\partial R_i} + \dfrac{\pi_j^i}{\pi^i} \dfrac{\partial x_j^*}{\partial R_i} \right)}{\pi^i} .$$

Defining the scale elasticity with respect to the rival's investment according to $e^{ij} = \dfrac{\pi_j^i x_j}{\pi^i}$ and making use of symmetry we can reformulate (22) as

$$(23) \qquad \frac{\partial \eta_i^*}{\partial R_i} = \frac{x_i^*}{\pi^i} \left[1 + \frac{R_i}{x_i^*} \frac{\partial x_i^*}{\partial R_i} - \frac{R_i}{x_i^*} \left(\frac{\partial x_i^*}{\partial R_j} e^{ii} + \frac{\partial x_j^*}{\partial R_i} e^{ij} \right) \right].$$

Making use of the interest rate elasticity of investment, $\varepsilon^* = -\dfrac{\partial x_i^* / x_i^*}{\partial R_i / R_i} > 0$, and (9) we define $\lambda = \dfrac{\partial x_j^* / \partial R_i}{\partial x_i^* / \partial R_i} = -\dfrac{V_{ji}^j}{V_{jj}^j}$ so that $0 < \lambda < 1$ if the investments are strategic complements and $e_j^{ii} \geq 0$, while $-1 < \lambda < 0$ if they are strategic substitutes and $e_j^{ii} \leq 0$. Making use of this notation we can express (23) according to

$$\frac{\partial \eta_i^*}{\partial R_i} = \frac{x_i^*}{\pi^i} \left[1 - \varepsilon^* (1 - (e^{ii} + \lambda e^{ij}))\right]. \qquad QED$$

Finnish Economic Papers – Volume 13 – Number 1 – Spring 2000

DO SHAREHOLDERS CARE ABOUT CORPORATE INVESTMENT RETURNS?*

OLLI HALTIA

European Investment Bank, 100, boulevard Konrad Adenauer, L-2950 Luxembourg

and

MIKKO LEPPÄMÄKI

Research Unit on Economic Structures and Growth, Department of Economics, P.O. Box 54, FIN-00014 University of Helsinki, Finland

This paper considers the apparent contradiction between the results of Artto (1997), who claims that shareholders in the paper industry have gained reasonable returns, and Pohjola (1996), who argues that a large shareholder value has been lost (the Artto–Pohjola Paradox). We show under fairly general conditions that the shareholders may enjoy reasonable return while the managers are simultaneously destroying value of the firm by allocating funds to bad investments. In this case the firm's shareholders suffer an opportunity loss equal to the value that could have been created if the firm had paid the funds out to them and they had invested the funds in equivalently risky projects. The paradox occurs if the growth rate of a firm's market value of equity is high enough to guarantee a non-negative return for the shareholders but too low in the sense that the shareholders would have earned more had the funds invested at the return (at least) equal to the opportunity cost of capital. Our results support Jensen's (1986) argument of the incentives of corporate managers to invest inefficiently, since here the shareholders do not necessarily challenge the management due to the fact that they may be perfectly happy and satisfied in financial terms. (JEL: D24, G31, L73)

1. Introduction

Do shareholders care about corporate investment returns? Is it possible that the shareholders are enjoying a reasonable return while the firm's managers are simultaneously destroying the value of the firm by making bad investments? It will be shown in the paper that this situation, although apparently contradictory, is indeed possible. In this case the firm's shareholders suffer an opportunity loss equal to the value that could have been created if the firm had paid the funds available for investment within the firm out to the shareholders and they

* *We thank Seppo Honkapohja, Michael C. Jensen, Erkki Koskela, Juha-Pekka Niinimäki, the editor and two anonymous referees of this journal for their comments. The views expressed are the sole responsibility of the authors, and do not necessarily reflect the view of the European Investment Bank (EIB).*

had invested the funds in equivalently risky projects.

This situation occurs especially if the shareholders judge the performance of the firm's managers only based on the shareholders' return whilst the losses due to the opportunity cost of capital are not transparent or observable to the shareholders. The result implies that the shareholders should not judge the performance of the managers only based on the shareholders' return but also based on the productivity of capital, which is directly controllable by the managers. In other words, the shareholders should ensure that the managers invest at the return at least equal to the opportunity cost of capital.

It is well known that the long-term economic growth depends crucially on the productivity of tangible and intangible capital. A typical feature for a development path of any economy has been a gradual move from the strategy of extensive growth, relying on expanding use of natural and other resources, towards the strategy of intensive growth, based on continuous improvement of efficiency in the use of existing production factors. Yet, as indicated by the recent study of McKinsey Global Institute (1996), there are huge differences in capital productivity even among the most developed nations across various industries.

In addition, capital productivity may vary drastically even between the firms in the same industry. Jensen (1993) argues that low productivity during the last two decades may have been caused by a failure of corporate internal control systems. Corporate control has been ineffective in dealing with a slow down in the markets, excess capacity and exit. It should be mentioned that our results further underline Jensen's original concern of the importance of the internal corporate control systems. While Jensen's (1986) proposition is based on the incentives of managers to invest inefficiently, our findings further support this argument. It will be shown in the paper that the shareholders do not necessarily challenge the management since they may be perfectly happy and satisfied in financial terms.[1]

The productivity and performance of the Finnish industry has recently been examined by Pohjola (1996). Applying the method established by Jensen (1993), Pohjola finds that in many industrial sectors, especially in the paper industry, the productivity of investments in 1986–1994 has been far too low from the point of view of the shareholder value generated. Based on Pohjola's findings one could conclude that the shareholders would have lost up to FIM 45 bn (some EUR 7.5 bn) over the period in terms of opportunity cost. In an apparent contradiction with Pohjola's findings, Artto (1997) shows that the shareholders in the paper industry should have enjoyed reasonable returns (around 15% IRR) over the period 1982–1996. Consequently, we have two views, one of which claims that a large shareholder value has been lost whereas the other concludes that shareholders have gained reasonable returns. This apparent contradiction will be called the *Artto–Pohjola paradox*.[2]

The remainder of this paper is organised as follows. In section 2 the opportunity costs of capital are modelled in terms of continuous time based on the so-called Jensen's method. Section 2 also derives the shareholders' return and relates that to the capital productivity measures. In section 3 we derive the sufficient condition for the Artto–Pohjola Paradox. Section 4 presents further analysis on the necessary conditions for the paradox. Section 5 concludes the paper.

2. *The opportunity cost of capital and the shareholders' return*

As argued by Jensen (1986), shareholders and corporate managers may have conflicting interests. In particular, the managers may have incentives to cause their firm to grow beyond the optimal size by realising investments, which do not yield positive net present values when

[1] *See also the recent interesting contribution by Kanniainen (2000) who considers corporate managers' inefficient investment behaviour under separation of ownership and control.*

[2] *Given the importance of the paper industry in the Finnish economy (the paper industry still accounts for about 25% of total exports), these controversial results have not only caused an academic debate but also been a source of public concern. See, for example,* the Helsingin Sanomat *19 March 1998.*

discounted at the relevant cost of capital. In other words, the shareholders would be better off if the managers would disgorge the cash to shareholders rather than invest it inefficiently. Thus a conflict may arise between the dividends to be paid out versus the funds to be invested within the firm. Following this line of thought, we employ Jensen's (1993) capital productivity measures and relate them to the explicit financial returns enjoyed by shareholders.

A firm can be considered as a value creating process for its shareholders and for the banking sector (financing the corporate debt), i.e. as a process creating value on the firm's capital structure. In general, a company generates shareholder value through two ways: (1) by a stream of dividends; and (2) by increasing the market value of the firm. In addition, a firm creates value for the banking industry in terms of paying the debt. Therefore, the total value V_T created by the firm up to date n can be expressed as follows:[3]

$$(1)\quad V_T = V_n + \int_0^n \delta_t e^{\rho(n-t)}dt + \int_0^n \beta_t e^{r(n-t)}dt,$$

where

- V_n = value of the firm at the future horizon date n (> 0);
- δ_t = payments to shareholders in the form dividends and net share repurchases (> 0);
- β_t = debt repayments including the principal and interest (> 0);
- ρ = the cost of equity;
- r = the riskless interest rate (riskless debt assumed).

The value of the firm at any moment of time equals the sum of the market values of equity and debt, i.e. $V_j = S_j + B_j$, j = 0, …,n . What is not paid out as a dividend or a repayment of debt from the corporate cash flow, is *invested within the firm*. To simplify our analysis we assume no taxes.[4] The ultimate aim of an investment (κ_t) can be taken as increasing the value of the firm, and thus contributing towards a higher shareholder value.

Consider next the firm that follows *an alternative investment strategy* of putting the funds κ_t available for investment (net of depreciation) to marketable securities (instead of investing in conventional R&D and machinery and equipment), yielding a return equal to their cost of capital, i, at the same risk level. Thus, i stands for opportunity return at the same risk level. Denote the value created by this alternative strategy as:

$$(2)\quad V'_T = V'_n + \int_0^n \delta_t e^{\rho(n-t)}dt + \int_0^n \beta_t e^{r(n-t)}dt + \int_0^n \kappa_t e^{i(n-t)}dt.$$

Based on assumed development of the equity value of the firm under the alternative investment strategy, a measure for capital productivity can be presented as follows.[5]

Measure J2. Under the alternative strategy, assume that the investment equal to depreciation suffices to maintain the intermediate cash flows ($C_t = \delta_t + \beta_t + \kappa_t$) but at the end of the period the firm has equity value of zero. Hence, $V'_n = 0 + B_n$.

$$(3)\quad J2 = V_T - V'_T = S_n - \int_0^n \kappa_t e^{i(n-t)}dt.$$

In above, J2 represents the difference between the equity value at the end of the period n (contributed to by investments made within the firm) and the future value of an alternative investment (e.g. marketable securities). Were J2 negative, the firm's investment strategy would

[3] *See Jensen (1993), which is based on discrete time. To simplify our analysis, we use a continuous time model.*

[4] *It is well known that with a tax regime which allows debt capital interest to be set off against tax liability, a company would be able to increase its expected annual net after-tax cash flow by gearing up. Thus, taxation would imply an incentive to change the firm's capital structure towards higher debt. The analysis to be presented in the following is based on comparison of two alternative investment strategies in a purely descriptive context. The effects of taxation would be the same for both investment scenarios and thus taxation should not influence the comparison outcome. A different but related question not addressed here is whether a tax regime* per se *changes the allocation of investment funds.*

[5] *Jensen (1993) presented four measures. Here only one measure will be reviewed but the analysis for the remaining measures should be straightforward following the argument to be presented in the following.*

not have been optimal in terms of creating value. A larger value could have been created if the firm had invested in alternative investment opportunities, e.g. marketable securities with the opportunity return i (at the same risk level).[6]

When looking at Jensen's capital productivity measures, it should be noted that the return on corporate investments is usually not observable to the shareholders, i.e. we have the case of *asymmetric information*. The shareholders evaluate the *explicit financial returns*, i.e. the share purchase costs vis-à-vis received dividends and capital gains in terms of increased share value. The total value for the shareholders measured at date n can be written as follows[7]:

$$(4)\quad R = S_n - S_0 e^{\rho n} + \int_0^n \delta_t e^{\rho(n-t)} dt,$$

where ρ is the cost of equity.

Does the shareholder's interest automatically guarantee that the firm's investment strategy yields a return on the investment equal to, or higher than, the opportunity cost of capital? In other words, does such a mechanism exist that corporate manager's bad investment decisions are necessarily reflected in the shareholder returns, leading possibly further to the management reorganisation by the owners? The answer to this question seems to be negative. The following discussion will elaborate that a shareholder may be perfectly satisfied even if the corporate investment strategy incurs a significant loss in opportunity cost terms.

[6] *For example, another measure presented in Jensen (1993), closely related to J2 measure examined in the present paper, is defined as follows: Assume the investment equal to depreciation is sufficient to maintain the initial equity value of the firm, but due to the allocation of (net) investment in the marketable securities the equity value does not increase. Hence,* $V'_n = S_0 + B_n$. $J1 = V_T - V'_T = S_n - S_0 - \int_0^n \kappa_t e^{i(n-t)} dt$. *Thus, J1 effectively compares the change of the equity value of the firm to the future value of investment in marketable securities.*

[7] *Artto (1997) used present value terms; here we apply future values. If R = 0, then ρ is the internal rate of return (IRR) on share purchases. If R > 0, then the shareholder enjoys higher return than IRR, i.e. IRR > ρ (the case of positive excess returns).*

3. *The Artto–Pohjola paradox*

In the following, we will demonstrate that the shareholders may enjoy a return equal to, or even higher than, their cost of equity (ρ) while they simultaneously suffer a loss in opportunity cost terms due to inefficient investments by corporate managers. In other words, the Artto–Pohjola paradox appears when the shareholders' return could have been increased had the investments within a firm yielded a return equal to the one available in alternative investment opportunities. For the sake of simplicity and without loss of generality in the rest of the paper, we assume constant dividends, $\delta_t = \delta$ and investments, $\kappa_t = \kappa$.

Consider now the relationship between the shareholders' return as defined in equation (4) and Jensen's capital productivity measure, equation (3). The value of end-period equity S_n is critical for the ultimate appearance of the Artto–Pohjola paradox. To show this, we assume that $S_n = S_0 e^{\gamma n}$, i.e. that the initial equity value grows to the end-period value at the constant rate γ. Recall that $J2 = S_0 e^{\gamma n} - \int_0^n \kappa e^{i(n-t)} dt$. After integrating, we get $J2 = S_0 e^{\gamma n} - (\kappa/i)(e^{in} - 1)$. The first term is the end-period value of equity, and the second term is the value of investments in marketable securities. The shareholder's total return is $R = S_0 e^{\gamma n} - S_0 e^{\rho n} + \int_0^n \delta e^{\rho(n-t)} dt$, and after integrating we get $R = S_0(e^{\gamma n} - e^{\rho n}) + (\delta/\rho)(e^{\rho n} - 1)$. The first term is the capital gain, and the second term is the received dividends.

Now we can find γ^J and γ^R such that J2 = 0 and R = 0, respectively. The *sufficient condition* for the Artto–Pohjola paradox is $R \geq 0$ and $J2 < 0$, which is valid when $\gamma^R \leq \gamma < \gamma^J$. After some calculation we get $\gamma^R = \ln[(\delta/\rho S_0)(1 - e^{n\rho}) + e^{n\rho}]/n$ and $\gamma^J = \ln[(\kappa/iS_0)(e^{in} - 1)]/n$. Now we can write the sufficient condition as follows:

Result 1: The Artto–Pohjola paradox appears if the growth rate γ of a firm's market value of equity is such that: $\rho + \ln[(\delta - \delta e^{n\rho})/(\rho S_0 e^{n\rho}) + 1]/n \leq \gamma < i + \ln[(\kappa e^{in} - \kappa)/(iS_0 e^{in})]/n$.

	$R \geq 0$	$R < 0$
$J2 \geq 0$	Investor Heaven	Black Hole
$J2 < 0$	Artto–Pohjola Paradox	Management Disaster

Figure 1: Jensen's Productivity Measure and Shareholders' Return.

Thus the Artto–Pohjola paradox appears when γ is high enough to guarantee a non-negative return ($R \geq 0$) for the shareholders but is too low in a sense that a larger shareholder value could have been created had the funds been paid out to the shareholders and invested further in equivalently risky projects. While Result 1 sets forth the general sufficient condition for the Artto–Pohjola paradox, an immediate and interesting further question is, in which kind of industries the paradox is likely. Answering the question properly would require empirical analysis of characteristics of various industries, which is beyond the scope of this paper.

Some further economic intuition can, however, be gained simply by inspecting the conditions that must hold to satisfy Result 1. Notice that the left-hand side of the inequality is smaller or equal to ρ when the term inside square brackets is strictly greater than zero but smaller than one. In order to demonstrate the relationship between the received dividends and the capital gain assume for a moment that $\delta = 0$. The non-negative shareholder return is possible even with zero dividends, but then the required γ has to be higher, in fact in this case it equals to ρ. Consider now the right-hand side of the inequality, and assume $\rho = i$. The strict inequality thus holds only when $(\kappa e^{in} - \kappa) > iS_0 e^{in}$, since only then $i \leq \gamma < i + \ln[(\kappa e^{in} - \kappa)/(iS_0 e^{in})]/n$. Assume now that dividends are positive, then we know that the required capital gain and thus γ, which guarantee the non-negative shareholder return, decrease. Therefore, lower returns on marketable securities (i) and/or smaller investments in marketable securities (κ) are needed for the appearance of the Artto–Pohjola paradox. In general, we can conclude that for a given growth rate of a firm's market value of equity the appearance of the Artto–Pohjola paradox is more likely for firms, which are intensive in investment compared to the dividends paid out.

The relationship between Jensen's productivity measure and the shareholders' financial return are further illustrated in Figure 1. For the sake of illustration, suppose that the shareholder IRR can be smaller than the cost of equity ρ, i.e. it is possible that $R < 0$ (negative excess returns). The Artto–Pohjola paradox appears in the low-left quadrant of the table, where the shareholders' return R is sufficient (positive or equal to zero) but Jensen's productivity measure is negative. Were the sufficient shareholder value R associated with high positive productivity figures, we would have the situation of "investor heaven" (the top-left quadrant). Third, low returns to the shareholder ($R < 0$), are likely to trigger a company restructuring ("management disaster"). Finally, the same scenario is likely for the top-right quadrant ("black hole"), where the firm absorbs equity while it does not invest.

4. *Analysis of the necessary conditions*

In order to gain further understanding on the necessary conditions for the paradox, we present some additional results. In particular, a consideration of the difference of R-J2 appears to generate some useful further results from an empirical viewpoint. Notice that positive R-J2 is a *necessary but not a sufficient condition* for the Artto–Pohjola paradox ($R \geq 0$): the positive difference R-J2 indicates that the financial return to the shareholder exceeds the capital productivity measure (furthermore with $J2 < 0$, the shareholder suffers an opportunity loss while being satisfied in financial terms). The analysis below indicates that the sufficient conditions for the Artto–Pohjola paradox are rather general. Note also that the magnitude of the difference R-J2 *per se* is not the main subject of our interest but the focus is instead on the sign of the difference. So, consider the difference:

$$\text{(5)}\quad R\text{-}J2 = S_n - S_0 e^{\rho n} + \int_0^n \delta e^{\rho(n-t)} dt - S_n + \int_0^n \kappa e^{i(n-t)} dt = \int_0^n \delta e^{\rho(n-t)} dt + \int_0^n \kappa e^{i(n-t)} dt - S_0 e^{\rho n}.$$

From above we get the following result:

Result 2: (R-J2) is independent on the end-period equity value S_n ($n > 0$).

In order to simplify our analysis we analyse the case where $\rho = i$, and then the difference becomes:

$$\text{(6)}\quad R\text{-}J2 = \int_0^n \delta e^{i(n-t)} dt + \int_0^n \kappa e^{i(n-t)} dt - S_0 e^{in}.$$

Two simple calculations are needed to derive the basic description for R-J2. First, the function is upward sloping with respect to time if its first derivative is positive. It appears that:

Result 3: $\partial(R\text{-}J2)/\partial n > 0$, if $\delta + \kappa > iS_0$ i.e. if the sum of annual dividends and investment is greater than the interest on the initial equity value.

The validity of this condition, of course, is an empirical matter. However, it should be noted that the condition is normally true if the company yields return on equity larger than the interest rate. It therefore is quite likely that the shareholder's return grows faster than does the productivity of capital J2. Second, we shall examine whether the difference R-J2 will become positive over time:

Result 4: $(R\text{-}J2) > 0$ if $\delta + \kappa > iS_0$ and n is greater than $n^* = [\ln(\delta + \kappa) - \ln(\delta + \kappa - iS_0)]/i$. Hence, (R-J2) will become positive with a sufficiently long time frame (n).

With regard to the above results, Figure 2 illustrates a possible outcome for the difference R-J2.[8] In brief, the shape of the curve R-J2 is determined by the fact that while the shareholders' opportunity cost increases over time only at the rate of cost of equity (ρ), the firm's opportunity cost of capital grows faster, at the rate of interest (i) plus the annual capital investment (κ).

Since the difference R-J2 is independent of the development of the equity value S_n – but the actual values of R and J2 are not – each value for S_n will generate different curves for R and J2. Therefore, with varying S_n, an infinite number of R and J2 curves can be generated, which all, however, will deliver the only one and the same difference curve R-J2.

Figure 3 below depicts one possible pair of R and J2, which generates the same difference curve R-J2 as in figure 2. In the figure, R is an increasing function of n whereas J2 is a decreasing function. The fact that R is increasing indicates that the cost of equity (ρ) is smaller than the shareholder IRR – indeed, the IRR here is 13%. The decreasing J2 curve indicates that the growth rate of the equity value (γ) is not large enough to compensate the opportunity cost of capital expenditure, determined by the available return when investing in alternative opportunities, e.g. marketable securities (an interest rate (i) as well as the cost of equity (ρ) of 10% was assumed in the calculations).

[8] *The figures are generated based on illustrative data assuming $\kappa = 0.25$ and $\delta = 0.025$ for periods 0–10 to be paid at the end of each period. Equity value S_0 of 1.4 is assumed throughout the period 0 and it is assumed thereafter to grow at 11% per year amounting to 4 at the end of the period 10. Investor IRR is 13% over periods 0–10. The cost of equity, ρ, and an interest rate, i, are both 10%.*

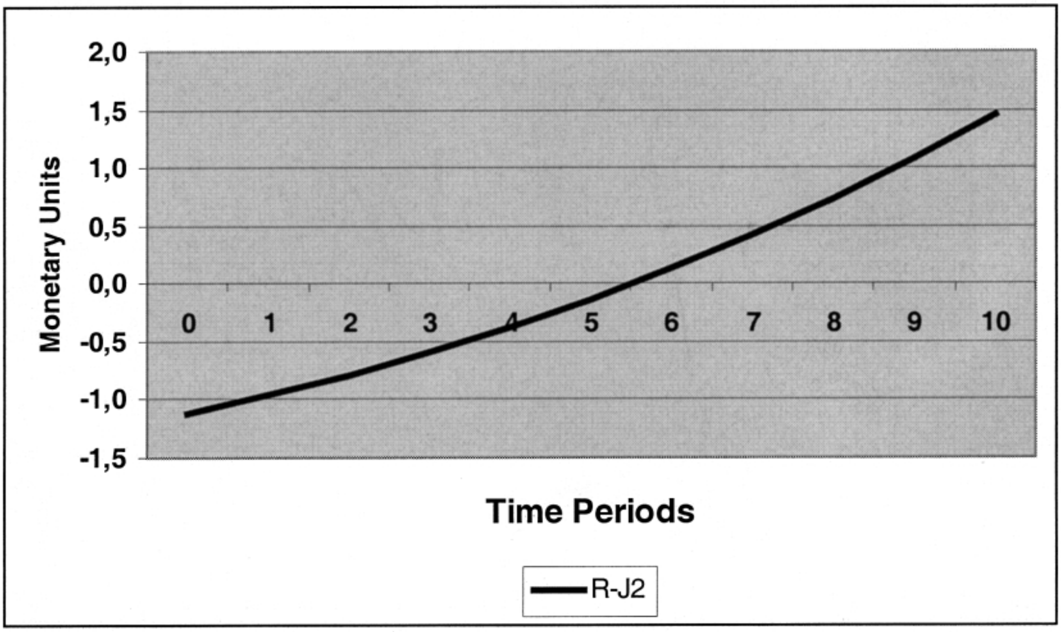

Figure 2: The Difference between Shareholder Return (R) and Capital Productivity (J2).

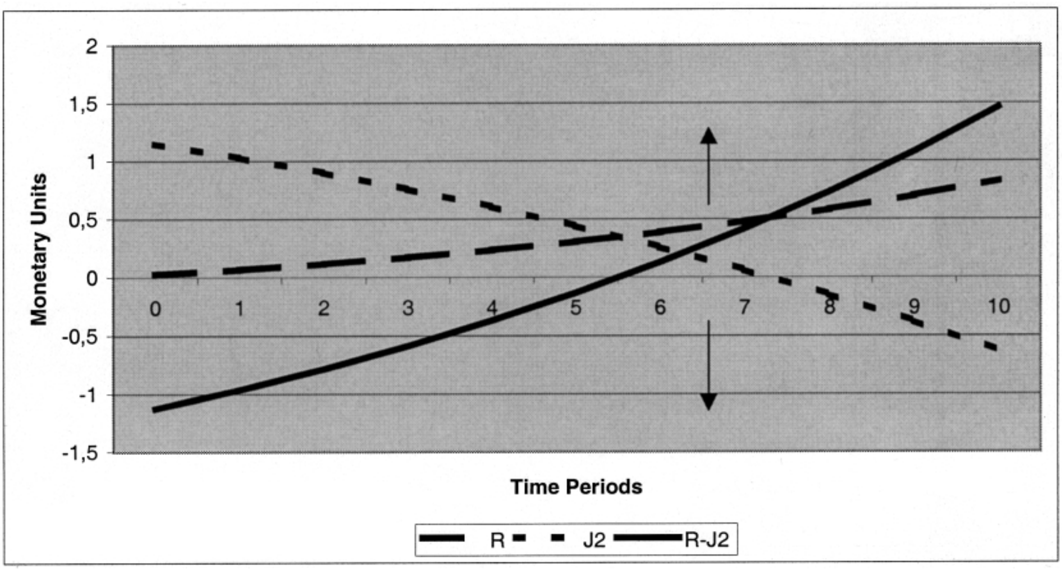

Figure 3: Shareholders' Return (R), Capital Productivity (J2) and the Difference (R-J2).

In figure 3, J2 reaches zero at the seventh period and thereafter J2 is negative whereas R remains positive and growing. Therefore, with n greater than seven, we have the Artto–Pohjola paradox with moderate shareholder return but too low equity value to compensate the lost return of investing in alternative opportunities than the firm itself. In general, a decrease in the end period equity value S_n would shift R and J2 downwards and an increase in the equity value would result in a shift upwards. However, the curve R-J2 is not affected by S_n, as was shown above.

5. Concluding remarks

It has been shown that, under quite general conditions, the shareholders' financial return R grows faster than Jensen's capital productivity (J2) with the lengthening of the time frame. In addition, with sufficiently long time frames R always becomes larger than J2. The Artto–Pohjola paradox appears if the firm's equity value increases at a relatively low rate. Here the opportunity cost of accumulating investment becomes large, resulting in negative capital productivity, measured in terms of J2. Simultaneously, the shareholders' financial return R may still remain positive indicating reasonably profitable investment in IRR terms (or in future value terms).

It is therefore possible that a firm goes on investing in low yielding assets while the shareholder still enjoys moderate returns without realising the opportunity losses arising from the investments. It can be concluded that the Artto–Pohjola paradox is not a real paradox. It should be emphasised that our results further underline Jensen's original concern of the importance of the internal corporate control systems. As argued by Jensen (1986), shareholders and corporate managers may have conflicting interests. In particular, the managers may have incentives to cause their firm to grow beyond the optimal size by realising investments, which do not yield positive net present values when discounted at the relevant cost of capital. While Jensen's proposition is based on the incentives of managers, our findings further support this argument. The shareholders do not necessarily challenge the management since they are perfectly happy and satisfied in financial terms.

It should also be noted that Jensen's productivity measures may indicate the same level of productivity for a highly investment-intensive company enjoying rapid increase in its market value of equity and, on the other hand, for a company with a rather stagnant equity value development and a low scale of investment activity. Even if the productivity measures are not directly observable *per se* for an investor, she typically gets worried if the share price of a capital-intensive firm performs poorly to some exchange index (i.e. average performance).

In general there are two possible reasons for the Artto–Pohjola paradox and for a low capital productivity rank *à la* Jensen. First, the investors may lack sufficient information in order to make a correct judgement on the expected returns of a firm's investment activity. In fact, companies do often have an incentive to conceal information on their investments, especially on R&D.[9] Therefore, the solution for improvement is in this case trivial: inform if you can.

The second possible explanation for a low capital productivity is that the firm's investment does not, indeed, yield sufficient returns (from a shareholder's viewpoint). The most straightforward solution for improvement is here to invest less and to pay more dividends. The scaled-down investment activity would immediately result in increased productivity (J2) whereas the increased dividend stream would simultaneously lift up the shareholder return (R). Also, if the firm's commitment to holding back its investments is considered convincing in the long term, the increased dividend stream may become capitalised in the market value of equity, lifting the productivity measures and the shareholder return further upwards. Alternatively, the firm might invest in buying back its own shares, hoping that it would result in a price increase. The increased capital gains would further improve the productivity performance as well as the shareholder return. Furthermore, the shareholder return could in this case still be improved by the fact that the dividend pool available per share held by the investors would grow. The result would thus be higher J2 and R, and

[9] *For example, Raision Yhtymä Ltd did not perform successfully in Pohjola's (1996) study since the research period did not capture the launching of a new product Benecol, after which the company's share price soared.*

a shift towards the "investor heaven" as indicated in figure 1.

The main message stemming from our analysis is that the simultaneous appearance of shareholders' reasonable return, on the one hand, and low capital productivity, on the other hand, is not necessarily paradoxical. This situation, although apparently contradictory, is possible when the growth rate of equity is within certain boundaries. In the long run, only firms, which are able to invest with returns higher than the opportunity cost of capital, are successful.

References

Artto, E.W. (1997). "Performance and International Competitiveness of Listed Paper Industry Groups 1982–1996 – Finland vs Sweden, Canada and USA." Helsinki School of Economics and Business Administration, Working Papers W-182.

Helsingin Sanomat (Jyrki Iivonen), (1998). "Paperiyhtiöiden investointien tuotosta syntyi kiivas oppiriita – talousprofessorien laskelmissa on kymmenien miljardien markkojen eroja." *Helsingin Sanomat*, *19.3.1998*.

Jensen, M.C. (1986). "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review* 76, 323–329.

– **(1993).** "The Modern Industrial Revolution, Exit and the Failure of Internal Control Systems." *The Journal of Finance* XLVIII, 831–880.

Kanniainen, V. (2000). "Empire Building by Corporate Managers. The Corporation as a Savings Instrument." *Journal of Economic Dynamics and Control* 24, 127–142.

McKinsey Global Institute (1996). *Capital Productivity*. Washington D.C.: McKinsey Global Institute.

Pohjola, M. (1996). *Tehoton pääoma*. Helsinki: WSOY.

PUBLIC CAPITAL AND PRIVATE SECTOR PRODUCTIVITY – A FINNISH PERSPECTIVE*

TOM BJÖRKROTH

Department of Economics and Statistics, Åbo Akademi University, Vänrikinkatu 3 B, FIN-20500 Turku, Finland

and

ANDERS KJELLMAN

Department of Economics and Statistics, Åbo Akademi University, Vänrikinkatu 3 B, FIN-20500 Turku, Finland

This paper focuses on the impact of the public sector capital on private sector productivity in Finland. The majority of previous contributions support the public capital hypothesis, i.e. that public capital plays an important role in enhancing private sector productivity. However, the contribution of Tatom (1991a) sheds new light on this issue. He shows that by correcting the previously used models of some of their most serious deficiencies, the impact of public capital is no longer significant for private sector productivity. We have applied Tatom's approach to Finnish data, and our results are similar to his findings. However, our results concerning the precedence between public capital and private sector output provide some evidence of causation running from public capital to private sector output, i.e. our results indicate that, if correctly targeted, public capital investment could affect private sector performance (JEL: E6, E62, H5).

1. Introduction

The public capital hypothesis which was suggested by Arrow and Kurz (1970) and has been developed by e.g. Ratner (1983), Aschauer (1989), Tatom (1991a, b) and Munnell (1992), states that the stock of public capital raises private sector output both directly and indirectly. The direct effect follows from the provision of intermediate services to the private sector.[1] The indirect effect arises from the complementarity of the flow of services from private and public

* *We are grateful for help received by Paul Wilkinson, Ilmo Pyyhtiä, Mikko Puhakka and for the valuable comments of two anonymous referees.*

[1] *An increase in the public capital stock implies then an increase in the marginal productivity of private capital. This should result in an increased demand for services of private capital goods and of private assets producing them, thus increasing output.*

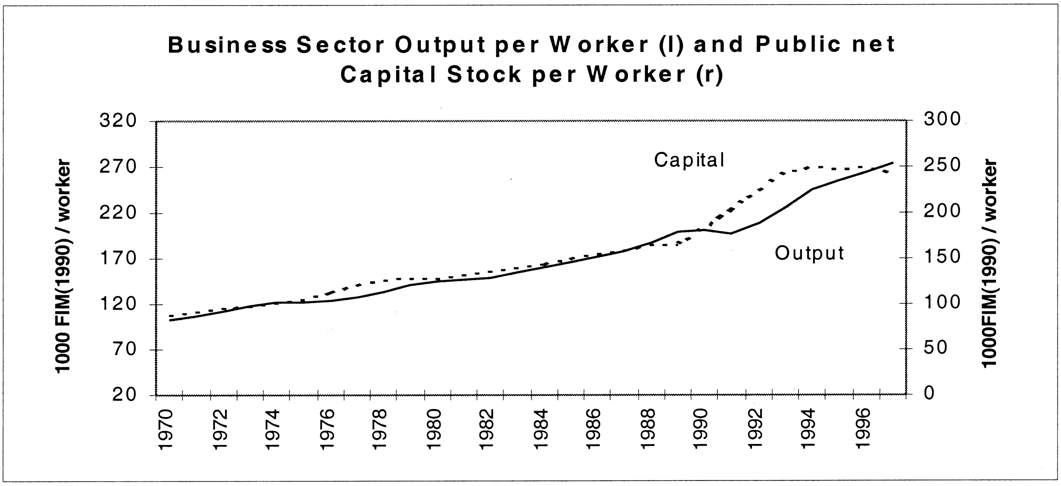

Figure 1: The relation between public capital stock (right scale) and private sector output (left scale) in Finland 1970–97. Source: Statistics Finland.

capital stocks, as noted by Munnell (1992).[2] Public capital investments will thereby increase the economic growth of a region.

Public capital relates to economic issues in a number of ways. However, let us first have a look at what public capital consists of. The infrastructure capital constitutes the most important part of the public capital formation in Finland, which includes roads, other networks of communications, military capital, educational and hospital facilities, power and water supply systems etc. One can hardly stress enough the importance of a well functioning infrastructure for the smooth functioning of an economy. To quote a metaphor from the World Bank (IBRD 1994 p.14): *"Infrastructure represents, if not the engine, then the 'wheels' of economic activity."* Perhaps the best way of recognizing the importance is to assess the effects of natural disasters that seriously damaged the infrastructure of an economy e.g. Honduras 1998. The quality of life and the productivity are also vulnerable to civil disturbances eroding the infrastructure. In this study we are not making any attempt to isolate the effects of the various components of the public capital stock on private sector performance. Instead the public capital is considered at the aggregate level.

The public capital hypothesis suggests that the stock of public capital increases the private sector output both directly and indirectly. Figure 1 shows that the business sector output per business sector worker, and public capital stock per business sector worker, are highly correlated (correlation coefficient of 0.98), which may lead one to conclude that there exists a causal relationship between these variables. The purpose of this paper is to investigate how the public sector capital formation really affects the private sector performance.

A common approach is to treat public capital as an input that enhances the productivity of both capital and labour. We will apply the approach suggested by Tatom (1991a) and try to avoid the statistical problems. However, emphasis will be put on the causal relationship between public capital and private capital. Is it so that you first need a strong private sector that can generate the resources needed for public investments, or will public investment spur private sector performance?

The paper proceeds as follows. In the next section we will have a look at previous studies and the criticism raised against these studies. In

[2] *It is a rather strong assumption that the flow of services from private and public capital are just complements. In some cases they may be substitutes as well. In that case an increase of public capital stock would tend to reduce private sector output.*

Table 1: Production function estimates of the output elasticity of public capital by level of geographic aggregation.

Author	Level of aggregation	Specification	Output elasticity of public capital
Holz-Eakin (1988)	National	Cobb-Douglas; Log levels	0.39
Aschauer (1989)	National	Cobb-Douglas; Log levels	0.39
Munnell (1990a)	National	Cobb-Douglas; Log levels	0.34
Ford & Poret (1991)	National	Cobb-Douglas; Log levels	0.39*
Tatom (1991b)	National	Cobb-Douglas Log levels	Not significant
Mamatzakis (1997)	National	Cobb-Douglas, Log levels	0.25
Costa et al (1987)	States	Translog; Levels	0.20
Eisner (1991)	States	Cobb-Douglas; Log levels	0.17
Mera (1973)	Japanese regions	Cobb-Douglas; Log levels	0.20
Munnell (1990b)	States	Cobb-Douglas; Log levels	0.15
Duffy-Deno and Eberts (1989)	Metropolitan areas	Log levels	0.08
Eberts (1986,1990)	Metropolitan areas	Translog; Levels	0.03

* The elasticity in United States. Ford & Poret (1991) study several other OECD countries.
Source: See Munnell (1992, p. 194) and Ford & Poret (1991), Tatom (1991b) for references.

Chapters 3 and 4 the data and model specifications will be addressed. The tentative results will be presented in Chapter 5 and the corrections for the main deficiencies in Chapter 6. Chapter 7 provides an indication of the direction of the causality between public capital stock and the private sector. Finally, the major conclusions are presented.

2. *Previous studies and major criticisms*

A number of studies on returns on infrastructure have been carried out. The results of some time series analysis show a strong positive relationship between spending on infrastructure and GDP growth. Much of the debate has its origins in Aschauer (1989) arguing that much of the decline in the US productivity that occurred in the 1970s was precipitated by declining rates of public capital investment. This view was shared by Munnell (1992). With reference to the Greek economy, Mamatzakis (1997) found support for the positive effect of the public capital stock on private sector productivity.[3] In a study of 26 Mexican three digit industries Shah (1992) finds evidence that public infrastructure has a small but positive multiplier effect on output. His results also indicate that in the long run there exists a small degree of complementarity in production between infrastructure and capital (pp. 32– 33).[4]

Lynde and Richmond (1992) (henceforth L&R) examine the effect of public capital on the production of aggregate output regarding the United States non-financial sector between the years 1958 and 1989. They find evidence to support the proposition that public capital infrastructure is of significant importance in the productivity of the private sector. L&R's (1992) findings correspond to those of Shah (1992), in that the services of public capital are more complementary than substitutes for services of private capital.[5]

Referring to several studies made at the national level, Munnell (1992) finds the impact of public capital to be too large to be credible. This is evident as the public capital sometimes has greater impact on output than private investments. On the other hand, she argues (Munnell, 1990b) that public capital has a positive impact on several measures of state level economic activity, such as output, investment and employment growth. Moreover, the coefficients of elasticity tend to be the same across studies, which is shown in Table 1.

[3] In fact, Mamatzakis focus is set on a broad definition of infrastructure, instead of measuring the effect of the entire public capital. (pp. 4–5).

[4] Shah (1990) uses the cost function approach as does Lynde & Richmond (1992).

[5] However, they do state that most of the public capital stock is investment in infrastructure facilities (p. 43).

Munnell (1992) describes the main features of the criticism concerning the general approach applied in previous studies. The criticism can be divided into the following four sub areas:

1. *Common trends in the output and public infrastructure data have led to spurious correlation.*

This criticism relies on the argument that these equations should be estimated using first differences. This is due to non-stationary data. The drift over time imposes the need to remove the trends. According to Munnell (1992), this leads to small effects of public capital investments. The effects are sometimes negative and generally not statistically significant.

2. *The wide range of estimates of the impact of public capital investments on output makes the empirical linkages fragile at best.*

Almost all studies report a positive and statistically significant output elasticity. To quote Munnell (1992, p. 193) this is: “,..., *amazing, as public capital spending contributes little to national output as conventionally measured.”* However, the contribution of Tatom (1991a) sheds new light on the issue. His results imply that, after addressing the main deficiencies of the earlier contributions, the public capital stock effect on private sector output is not statistically different from zero.

3. *The direction of causation may run from high levels of output to greater public capital investments, rather than the other way around.*

Munnell (1992) finds this criticism legitimate, but adds that further capital investments, both private and public, goes hand in hand with economic activity.[6] However, this mutual influence may exist without tainting the coefficient on public capital. In their study, Eberts and Fogarty (1987) examined the problem of causality, using public and private investments data from 1904 to the year 1978 for 40 metropolitan areas. The results implied the existence of a causation running in two directions. Public capital investments affected private investments mainly in cities that experienced most of their growth in the1950s. In southern cities that grew faster since the 1950s, the causation ran from private to public capital investment.

Aschauer (1989) tried to solve this problem of causality by: a) using lagged infrastructure investment as an instrument for contemporaneous investments in his regressions, b) splitting infrastructure investments into components judged *ex ante* to be important and those not important.

Ford and Poret (1991) criticise this Aschauer’s method, especially regarding the use of lagged infrastructure variables. They find it unlikely that this technique could successfully deal with simultaneity bias. On the other hand, Ford and Poret give Aschauer some credit for his splitting of the infrastructure group. Especially as the group considered important ex ante was found more important in explaining the deceleration in total factor productivity (TFP). Moreover, they conclude (p. 65) that: *“This is a more convincing control for reverse causality than the use of instrumental variables.”* Despite the criticism, Ford and Poret (1991) carry on their own analysis using the Aschauer approach on a broader range of data, still aware that the problem of reverse causation is hard to overcome.

4. *The production function framework is inadequate.*

The critics of the production function approach argue that the omission of input prices in the production function framework will bias the estimated coefficients. It is also argued that this framework involves too many restrictions on the technology and behaviour of firms. As a solution to this problem, the critics suggest that one should estimate cost functions, which allow one to disentangle the effects of infrastructure investments. As reported by Munnell (1992), the cost function approach has been adopted by

[6] *Compare with the results of Blomström et al. (1993). They find that economic growth induces capital formation more than capital formation induces growth.*

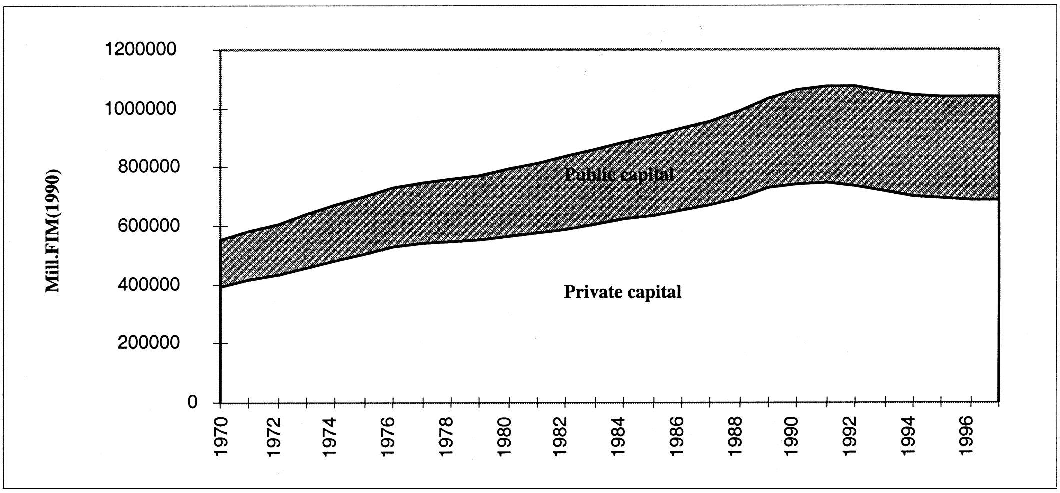

Figure 2: The net public and private non-residential capital stock in Finland between 1970–1997 in constant 1990 prices. Source: Statistics Finland.

a few authors. However, in this study we will proceed along the lines of Tatom (1991a). We seek to apply this methodology on a set of Finnish data during the period 1970–1997. Furthermore, we wish to focus on the issue of direction of causation.

3. The data

The data has mainly been supplied by Statistics Finland. The variables that we are using are the following:

- Q_t = Business sector output in FIM (1990) Million
- K_t = The net non-residential stock of private capital at year (t) in FIM (1990) Million.
- KG_t = The public net non-residential capital stock at year (t) in FIM (1990) Million.
- h_t = Business sector hours in year (t) expressed in units of 100000 hours.
- t = a time trend (t =1,...,28), which also can be seen as capturing the technological development
- c_t = capacity utilisation rate[7] of the manufacturing industry at each year (t)
- k_t = $c_t{*}K_t$, where c_t is the capacity utilisation rate

The net public and private non-residential capital stock between 1970–1997 is presented in Figure 2.

Figure 2. shows that the total net non-residential capital stock has grown steadily until years 1989–1990. After this, the private net capital stock has declined.

[7] The capacity utilisation rate (c) was provided by the Bank of Finland and Statistics Finland and the values for the aggregate manufacturing industry are used in the calculations. The values for the period 1975 to 1992 were calculated as annual averages from observations made twice a year by Bank of Finland. The values from 1993 to 1997 were calculated as annual averages from monthly data, provided by both Bank of Finland and Statistics Finland. The capacity utilisation rates for years 1970 to 1974 had to be estimated. Ford and Poret (1991, p. 84) measure the capacity utilisation rate by using the percentage of firms operating at full capacity. This variable has been indirectly used to estimate the capacity utilisation rate used in this study. The relation between the percentage of firms operating at full capacity and the capacity utilisation rate is logarithmic, but using e.g. 10 + c^*_i/100 for the base year 1975 and using it as an index to estimate the measured capacity utilisation rate, yields satisfactory results (c is F&P's, 1991, measure of capacity utilisation for Finland and subindex i denotes year in question). The correlation between the actual series and the one predicted by the formula above is 0.996 for the period 1975 to 1997 and even higher for some subperiods. However, this new series tends to slightly overestimate the fluctuations in the observed capacity utilisation rate, but should not lead us too far astray.*

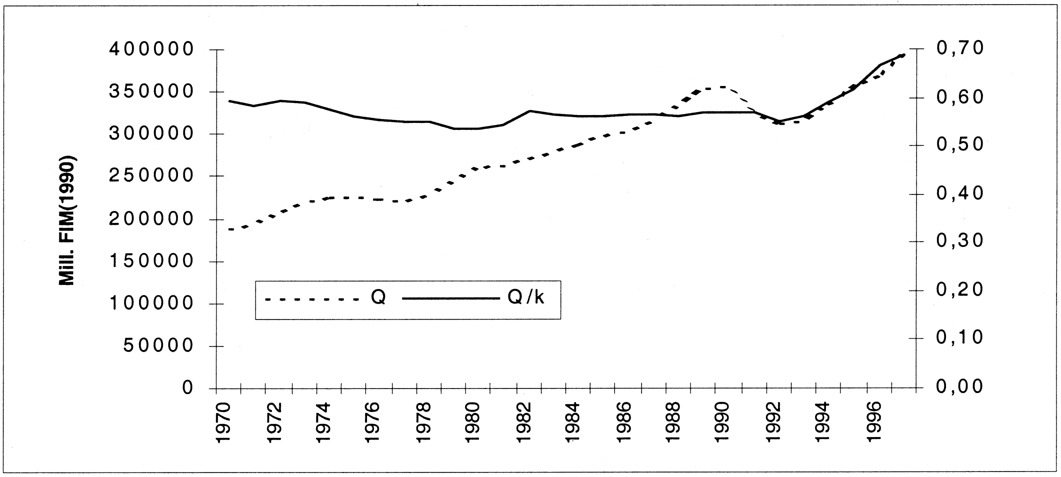

Figure 3: Business sector output, Q_t (left axis) in FIM (1990) and its ratio to the flow of services from net private non-residential capital stock, Q_t/k_t (right axis).

Figure 3: Constant FIM (1990) business sector output, Q_t (left axis) and its ratio to the flow of services from net private non-residential capital stock, Q_t/k_t (right axis)

Figure 3 shows the business sector output and the impact of the changes in net non-residential private capital stock. This ratio, which is also the dependent variable in regressions below, reveals that the slowdowns, especially in the mid-1970s but also in the early 1990s, have seriously affected private sector productivity. Whether the reason to these slowdowns is the inappropriate or excess formation of public capital will be answered below.

4. The model

The conventional model, focusing solely on infrastructure[8], is similar to that of F&P (1991). A Cobb-Douglas technology is assumed to produce private sector output (Q), using private sector inputs (PIN) and infrastructure capital (INF). Denoting this in logs:

(1) $q = a + b \cdot \text{inf} + c \cdot \text{pin}$

[8] *See Madden (1998) for the effect of telecommunications investment on economic growth.*

Since the total factor productivity (TFP) can be measured as a difference of the growth rate of the output (q) and of the private inputs (pin) we may write:

(2) $\text{TFP} = q - \text{pin} = a + b\text{inf} + (c - 1)\text{pin}$

This is equivalent to equation (1.7) in Aschauer (1989, p. 183), which is used in estimations. The parameter estimates should be interpreted as the following:

b = The elasticity of output(and TFP) with respect to infrastructure
c = The elasticity of output with respect to the private-sector input bundle
c = 1 implies that there are constant returns to scale in the private-sector and thus pin drops out of Equation 4.2.

According to Tatom (1991a), the essential shortcomings of models similar to this are:

a) The omission of a trend
b) That they contain non-stationary variables
c) The ignorance of the effect of relative price of energy on productivity

The shortcomings a) and b) will be dealt with in this study. The authors are aware that the omission of energy price effects on productivi-

Table 2: The results of the estimations.

Specification	(1)*		(2)*		(3)	
Variable	*estimate*	*t-ratio*	*estimate*	*t-ratio*	*estimate*	*t-ratio*
Constant	1.514	8.69	0.946	5.10	0.980	4.76
$\ln \frac{KG_t}{K_t}$	0.365	2.82	–0.066	–0.64	–0.03	–0.26
$\ln \frac{h_t}{k_t}$	0.793	9.38	0.700	9.06	0.699	7.56
t	0.028	7.28	0.017	4.49	0.017	3.50
t^2			0.0004	5.40	0.0004	4.54
Rho (1)					0.778	5.27
Rho (2)					–0.624	–4.22
F	48.22		57.25			
R^2 adj	0.84		0.89		0.96	
DW	0.74		0.85		1.56	
Jarque-Bera	1.978		1.136		7.435	

* Specifications (1) and (2) were estimated using a heteroskedasticity consistent covariance matrix.

ty may bias the coefficients. There exists a certain risk that possible productivity losses, due to changes in the price of energy, can be falsely attributed to the changes in public capital.[9] According to his critics on the previous contributions, Tatom (1991a) starts with a slightly different model, which we are going to apply:

$$(3)\quad Q = Ah_t^{\alpha} k_t^{\beta} kg_t^{\delta} e^{(rt+\varepsilon_t)}$$

with A representing a scale parameter, h_t denoting business sector hours. k_t is the flow from services from K_t , the constant FIM net non-residential stock of private capital at the end of the previous year. kg_t is the flow of services from KG_t , the net non-residential public capital stock at year t. r is the rate of disembodied technical change and t is the time trend. ε_t is a normally and independently distributed random disturbance term.[10] Assuming constant returns to scale[11] allows us to write:

$$(4)\quad \ln(Q_t/k_t) = \ln A + \alpha\ln(h_t/k_t) + \delta\ln(KG_t/K_t) + rt + \varepsilon_t$$

5. Estimation results

Table 2 shows three estimation results. Model Specification (1) corresponds to Equation (4). In Specification (2) a quadratic time trend is added[12] and under the assumption that evidence of serially correlated errors reflect truly autocorrelated residuals, this model is further corrected for serial correlation and these results are shown as specification (3). Specifications (1) and (2) are estimated using the White's heteroskedastic-consistent covariance matrix.

A specification similar to that of (2) was estimated without the heteroskedastic consistent

[9] See Tatom (1991b, p. 7–8) , especially footnote 13, for an explanation on this.

[10] The same capacity utilisation rate, c_t, is applied for both public and private capital i.e. $k_t = c_tK_t$ and $kg_t = c_tKG_t$. The assumption that the capacity utilisation rate is the same for private and public capital is made due to lack of sufficient data of public sector utilisation. We are aware that the capacity utilisation level in the public sector could be lower than in the private sector. However the capacity utilisation rates do not need to be equal for the analysis to proceed. Such a scenario then requires an assumption that the relation between these utilisation rates is constant over time. In the equations used a proportionality assumption of public and private utilisation rates would lead to a situation where the proportionality is included in the constant term (A).

[11] I.e. $\alpha + \beta + \delta = 1$, implying that a proportional rise in each resource raises private output (Q) by the same proportion.

[12] Tatom (1991b, p. 7) argues that omission of a significant time trend can bias the the coefficients and standard errors of the estimates, which is verified in ourthe estimations below.

covariance matrix (*not shown* (2')), and it formally passed the diagnostic checks. The cumulative sum of recursive residuals stays well inside the boundaries at the 5% level, as does the cumulative sum of squared recursive residuals. The recursive residuals (rr) show two sets of positive values. Firstly, between observations 11 (year 1980) and 14 (year 1983), and secondly, between years 1994 and 1997. At years 1991 and 1992 there are two negative rr-values.[13] This might reflect the difficulty in catching the more dramatic fluctuations in the dependent variable at the end of the period. Heteroskedasticity does not seem to be a problem with this specification. The value of Harvey's (1982) Equation 2.12 ~ F(7,7) is 2.9446 which does not exceed the critical value at the 5% level (3.787) The heteroskedasticity test where the errors were regressed on the dependent variable suggests homoskedasticity, but the Breusch-Pagan-Godfrey test statistic survives at the 5% level but not at the 10% level. This also holds for the ARCH-test. Harvey's test survives only the 1% level and the Glejser-test suggests homoskedasticity only at the 0.5% level.

These results suggest that the possible heteroskedasticity is associated with one of the independent variables. Having plotted the residuals from Specification (2') against the independent variables[14], no evident relationship between error variance and level of independent variables can be found. We tried to run a weighted regression using $1/\ln(KG_t/K_t)$ as a weight, but this did not significantly alter the coefficients. Therefore, the White's heteroskedastic-consistent covariance matrix was used in estimating (2). It was also obvious from the plot of residuals from Specification (2), that (2) suffered from serially correlated errors (see Figure A6 in the Appendix).

In specification (3) we have corrected for a second order autocorrelation (AR(2)) in the error term. It should be noted that the coefficient for public capital is not significant in specification (2) so this correction could at least improve its parameter estimate. The AR(2) model for the error term results in a stationary model, but in complex roots the autoregressive process displays pseudo periodic behaviour with a damped sine wave. This correction does not materially alter the coefficient estimates of specification (2). We also tried to remove the significant quadratic time trend from Specification (3). However, this only verified the concern of Tatom (1991a) that excluding a significant time trend from a model may bias the estimated coefficients, especially those of the variables which are correlated with the omitted time trends.

The estimation results imply that the effect of the aggregate net non-residential public capital is not significantly different from zero. The significant parameter estimate of t reveals that the effect of the disembodied technical change has been significantly positive. The significant quadratic time trend implies that the productivity (growth) is accelerating.

6. *Test of stationarity, further estimation results and co-integration?*

The stationarity of variables removes the possibility of getting spurious results from the regressions.

Test equation for the levels of each variable Y is:

$$(5)\quad \Delta Y_t = \alpha + \beta Y_{t-1} + \delta t + \sum_{j=1}^{p} \gamma_j \Delta Y_{t-p} + \varepsilon_t,$$

where p was automatically set by the SHAZAM software and is reported in Table 6.1.

The relevant statistic for this unit root test is the t-statistic for the lagged level of the variable. If β is significantly less than zero the variable in question is stationary. For this sample size (N = 28) with undifferenced variables, the asymptotical critical value of the t-statistic for β is about –2.95 when the time trend is not included, and about –3.58 when the trend is present (5% level of significance).[15] The criti-

[13] *See Figures A.1 and A.2 in the Appendix.*

[14] *See Figures A.3 to A5 in the Appendix.*

[15] *These t-values were calculated using the Table 7.2. in Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1992) Co-integration, Error Correction, and Econometric Analysis of Non-stationary Data. The critical val-*

Table 3: The results from testing for non-stationarity.

Variable	t-statistic	critical value (5%-level)	z-test	crit-value (10%-level)*	p
$\ln \frac{q_t}{k_t}$	–0.470[a]	–2.95			1
$(1–B) \ln \frac{q_t}{k_t}$	–2.86[a]	–2.97	–13.24	–11.2	0
$(1–B)^2 \ln \frac{q_t}{k_t}$	–4.09[a]	–2.98			1
$\ln \frac{KG_t}{K_t}$	1.46[a]	–2.95	2.08	–11.2	0
$(1–B) \ln \frac{KG_t}{K_t}$	–2.31[a]	–2,97			1
$(1–B)^2 \ln \frac{KG_t}{K_t}$	–3.51[a]	–2.98			2
$\ln \frac{h_t}{k_t}$	–2.59[a]	–2.95	–1.36	–11.2	0
$(1–B) \ln \frac{h_t}{k_t}$	–2,48[a]	–2.97			2
$(1–B)^2 \ln \frac{h_t}{k_t}$	–4.78[a]	–2.98			2

* This critical value was given by the statistical software. a=*constant, no trend.*

cal values for the once and twice differenced variables were calculated after dropping the undefined observations, which caused only minor adjustments.

The results in Table 3 above indicate that all the variables above have to be differenced twice in order to achieve stationarity. In Section 7 these stationary variables were used in a new estimation of the original model

Further estimation results

Table 4 shows the estimation results using stationary transformations of the independent variables. Since the variables were differenced twice, the sample period was adjusted to remove undefined observations. These specifications are also estimated using White's heteroskedastic-consistent covariance matrix.

ues for the estimate of β at 1% and 10% levels, using a constant without a trend were about –3.685 and –2.62 respectively. Including a time trend yielded the critical values of about –3.224 (10% level) and –4.323 (1% level).

Specification (4) above reaches a rather modest adjusted R^2 and has, according to the Jarque-Bera statistic, normally distributed residuals.

Table 4: The new estimation results.

Specification	(4)		(5)	
Variable	*estimate*	*t-ratio*	*estimate*	*t-ratio*
Constant	0.0004	0.05	0.0003	0.07
$(1–B)^2 \ln \frac{KG_t}{K_t}$	–0.034	–0.16	–0.004	–0.03
$(1–B)^2 \ln \frac{h_t}{k_t}$	0.529[a]	5.90	0.507[a]	7.25
$(1–B)^2 \ln \frac{Q_{t–1}}{k_{t–1}}$			–0.32[c]	–1.92
$(1–B)^2 \ln \frac{h_{t–1}}{k_{t–1}}$			0.092[b]	2.35
t	$0.11 \cdot 10^{–4}$	0.02		
F	7.706			8.579
R^2 adj	0.45			0.55
DW	2.23		Durbin's h	0.844
Jarque-Bera	0.861			0.609

a, b and c indicate significant estimates at the 1%, 5% and 10% level.

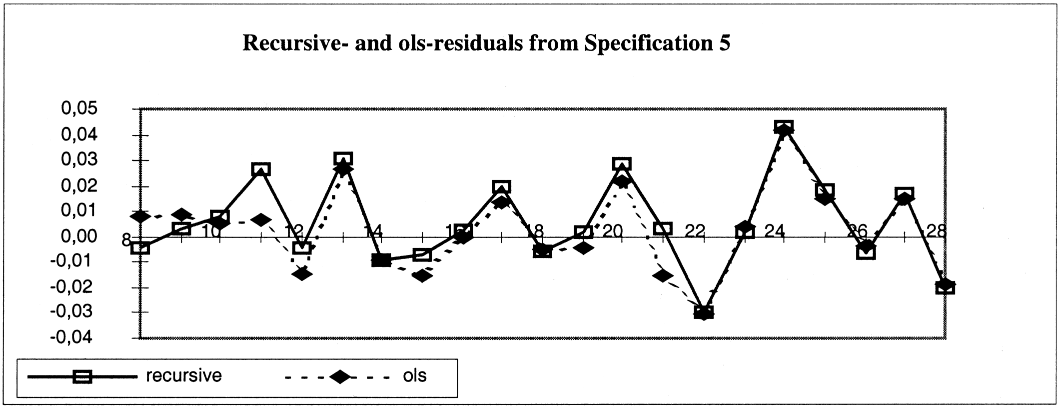

Figure 4: Test of misspecification with ols- and recursive residuals.

The result regarding serially correlated errors suggests that the errors are not serially correlated. Therefore, the estimates can be judged to be rather efficient and consistent with Tatom's (1991a, Equation. 7, p.11) results. The ratio of business sector hours to flow of services from private capital clearly dominates the model, while leaving no significant effect of the public capital at the 5% level of significance. The time trend is not significantly different from zero. In Specification 5 we have added two lagged variables; the dependent variable and the ratio of business sector hours to flow of services from private capital. This was performed in order to further lessen the possibility of serial correlation in the error term resulting from an autoregressive process. The additional explanatory variables improve the adjusted R^2. Durbin's h statistic and Jarque-Bera indicate that there is neither evidence of serially correlated residuals nor of inefficient parameter estimates. Again, there is some evidence that aggregate public capital has no significant role to play in determining the private sector output. Adding a (statistically not significant) time trend to Specification 5 did not dramatically alter the results. The effect of public capital stock still remained highly insignificant (t-statistic = 0.02). This was also the case when only t^2 was added.

The tests for heteroskedasticity and Ramsey's specification error test (RESET) do not suggest any problems of either heteroskedasticity or misspecification. In Figure 4 we have plotted the recursive residuals together with the ols-residuals. The recursive residuals seem to follow the ols-residuals well enough to remove any doubts of misspecification. The cusum and cusumsq tests (Tables A7 and A8 in the Appendix) show that both model specifications also pass this formal test.

Are the variables co-integrated?

Model specifications (1) to (3) could be applied in this context provided the variables used are co-integrated. However, the results from cointegration regressions and from the Dickey-Fuller tests on the residuals reveal that the undifferenced variables are not co-integrated, which also holds for the first differences. The first differences provide significant t-test-statistics, but the z-statistics are still too small in absolute values to allow one to reject the null hypothesis of a unit root. Moreover, the condition of a high R^2 statistic and a Durbin-Watson statistic close to 2 is met neither by the levels nor first differences of the variables.

Table 5 clearly indicates that the estimates in models (1) to (3) are results of spurious correlations and that the results from Specifications (4) and (5) are more reliable. Due to the better performance and as a result of the diagnostic checks of Specification (5) we argue that it is to be preferred to the more parsimonious Specification (4).

Table 5: Test of co-integration of levels and first differences.

Variables in	Test-statistic	Critical value (10%)	R^2	DW	AIC	SC
Levels						
Cointegrating regression: Constant, no trend			0.68	0.67		
Dickey-Fuller on residuals						
No constant,	z = –7.68	–22.7			–7.241	–7.192
no trend	t = –1.79	–3.45				
Cointegrating regression: Constant, trend			0.88	0.81		
Dickey-Fuller on residuals						
No constant,	t = –6.36	–3.45			–8.147	–7.998
no trend						
First Difference						
Cointegrating regression: Constant, no trend			0.61	1.21		
Dickey-Fuller on residuals						
No constant,	z = –16.62	–22.7			–8.353	–8.304
no trend	t = –3.72	–3.45				
Cointegrating regression: Constant, trend			0.66	1.30		
Dickey-Fuller on residuals						
No constant,	z = –19.28	–28.5			–8.576	–8.528
no trend	t = –4.48	–3.84				

7. The problem of precedence

The framework developed by Hsiao (1981), formerly applied by Madden et al. (1998), was used to test for direction of causation between the public capital stock and private sector output.[16] The summary of the results from testing the precedence between the ratio of public to private capital and private sector productivity are reported in Table 8. Table 6 shows the Akaike (1969) Final Prediction Error (FPE) of the one-dimensional autoregressive processes for the two variables being tested. At this stage only four lags were considered, because of the small sample size. Due to the differencing and the lag structure, initial observations were dropped in order to avoid undefined observations. The results show that one lag is enough to minimise the FPE for $ln(Q_t/k_t)$, while two lags are required for $ln(KG_t/K_t)$. These lags correspond to *m* lags for variable *y* in the contribution of Hsiao (1981).

The next step is to regress the variable *x* on *y*, with an optimal lag length *n*, which is performed in Table 7. Treating $ln(Q_t/k_t)$ as y, suggests that four lags of $ln(KG_t/K_t)$ are needed to minimise FPE. Only one lag is needed when $ln(KG_t/K_t)$ is treated as the dependent variable.

The FPE statistic from the unrestricted regressions is then used to determine whether there is at least weak evidence of precedence running from *x* to *y*. If the FPE is larger in the

Table 6: The Akaike (1969) FPEs of one-dimensional AR-processes for twice differenced $\ln(KG_t/K_t)$ and $\ln(Q_t/k_t)$ following Hsiao (1981).

Order of lags	FPE ($*10^{-3}$) of $(1–B)^2 \ln Q_t/k_t$	FPE($*10^{-3}$) of $(1–B)^2 KG_t/K_t$
1	***0.75374***	0.64635
2	0.77088	***0.59879***
3	0.77099	0.65510
4	0.81100	0.71889

[16] *The test procedure was as follows: 1) To test whether x precedes y, y was regressed on $y_{t–m}$. The optimal lag length m was determined by minimising Akaike's (1969) Final Prediction Error-criterion (FPE) for 1≥ m ≥ 4, 2) The variable y_t is regressed on m lags of itself and $x_{t–n}$. The optimal lag length of x_t is again determined by minimising FPE as 1≥ n ≥ 4, 3) The FPEs of steps 1) and 2) are compared. If $FPE_1 > FPE_2$, there is weak evidence that x precedes y ($x \rightarrow y$), 4) an F-test was used in step 2) to test whether the estimates of x:s are jointly equal to zero. If this is rejected there is weak evidence that x precedes y. and 5) If the conditions in both step 3) and 4) are satisfied, there is strong evidence that x precedes y, i. e. $x \Rightarrow y$.*

Table 7: The optimal number of lags of the independent variables.

Controlled variable	Manipulated variable	Optimum lag of manipulated variable	FPE (*10^{-3})
$(1-B)^2 \ln Q_t/k_t$	$(1-B)^2 KG_t/K_t$	4	0.68590
$(1-B)^2 \ln KG_t/K_t$	$(1-B)^2 \ln Q_t/k_t$	1	0.70471

unrestricted regression, there is weak evidence of this 'causation'. If a joint F-test on the added variables simultaneously exceeds the critical values of the F-distribution, there is strong evidence that *x* precedes *y*. In Table 8 a summary of the test results is shown.

The regressions were estimated using White's heteroskedastic consistent covariance matrix. Only the adjusted R^2 together with Durbin-Watson (D-W) and Jarque-Bera (J-B) of the diagnostic statistics are shown here in order to check for the fit, serial correlation and normality of residuals. The decisive statistics, FPE and the joint F-statistic are also reported. The results provide strong evidence at the 10% level of significance, that $ln(KG_t/K_t)$ does precede $ln(Q_t/k_t)$. At the 5% level there is only weak evidence left of this precedence since FPE_1 is greater than FPE_2. There is, however, no evidence that the variable $ln(Q_t/k_t)$ would cause changes in $ln(KG_t/K_t)$. The result of precedence is important even if there was no significant effect of the $ln(KG_t/K_t)$ on private sector output in the estimation stage. This refers to the results of Easterly and Rebelo (1993) that particular investments, especially in infrastructure and transport, are consistently, and positively, correlated with economic growth. Easterly and Rebelo also discuss the problem of causation. They argue that if the direction of causation was from growth to public investment, all types of public investment should be associated with growth, but (to quote, p 433): *"One piece of indirect evidence against reverse causation is that only transport and communications investment and general government investment are robustly correlated with growth."*

8. Conclusions

Despite the fact that many contributions affirm that increased public capital investment enhances private sector productivity and increases output, our results suggest that this is not necessarily true from a Finnish perspective. We have studied Finnish data from 1970 to 1997, and found evidence that changes in the aggregated net non-residential public capital stock has no significant effect on private sector productivity. These findings are consistent with those of Tatom (1991a) but ought not to be seen

Table 8: Summary of results from testing for precedence between $\ln(Q_t/k_t)$ and $\ln(KG_t/K_t)^*$.

The ratio of public net non-residential capital stock to private net non-residential capital stock (n=22)	R^2-adj.	D-W	J-B	FPE	F	S
$\Delta^2 \ln(Q_t/k_t) = 0.005–0.382\Delta^2 \ln(Q_{t-1}/k_{t-1})$	0.08	1.95	0.337	0.754		(1)
$\Delta^2 \ln(Q_t/k_t) = 0.002–0.40\Delta^2 \ln(Q_{t-1}/k_{t-1}) + 0.09\Delta^2 \ln(KG_{t-1}/K_{t-1}) + 0.02\Delta^2 \ln(KG_{t-2}/K_{t-2}) + 0.57\Delta^2 \ln(KG_{t-3}/K_{t-3}) + 0.68\Delta^2 \ln(KG_{t-4}/K_{t-4})$	0.28	2.00	1.186	0.686	2.561^c	(2)
$\Delta^2 \ln(KG_t/K_t)= 0.003–0.77\Delta^2 \ln(KG_{t-1}/K_{t-1}) –0.29\Delta^2 \ln(KG_{t-2}/K_{t-2})$	0.35	2.00	32.96	0.599		(3)
$D^2 \ln(KG_t/K_t)= – 0.002–0.76\Delta^2 \ln(KG_{t-1}/K_{t-1}) –0.29\Delta^2 \ln(KG_{t-2}/K_{t-2}) + 0.87\Delta^2 \ln(Q_{t-1}/k_{t-1})$	0.32	1.94	29.6	0.705	0.158	(4)
Conclusion	**$\ln(KG_t/K_t) \Rightarrow \ln(Q_t/k_t)$, and $\ln(Q_t/k_t)$ does not precede $\ln(KG_t/K_t)$**					

* $=\Delta^2$ is equivalent to $(1-B)^2$, c refers to significance at the 10% level and S indicates the specification.

as diametrically opposite to those contributions, which focus only on some specific parts of the capital stocks.

We tested for precedence and found evidence, not particularly strong, however, that changes in the aggregated level of public capital stock precede private sector performance. This is consistent with at least the arguments of Easterly and Rebelo (1993), IBRD (1994) and Mamatzakis (1997), and has some implications for advocating public investment in sectors that are positively correlated with economic growth, e.g. transport and communication. The fact that many factors affecting private sector productivity have not been used, or even required, to explain most of its variations does not mean we consider them irrelevant. We are fully aware that the addition of energy prices, for example, to the models may alter the parameter estimates. The present lack of available data puts this task, together with the break-down of the public capital stock, on the agenda of further research.

References

Arrow, K.J., and M. Kurz (1970). *Public Investment, the Rate of Return, and Optimal Fiscal Policy*. Baltimore: John Hopkins Press.

Aschauer, D. A. (1989). "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23, 177–200.

Banerjee, A., J. J. Dolado, J.W. Galbraith, and D. Hendry (1993). *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. New York: Oxford University Press.

Blomström, M., R.E. Lipsey, and M. Zejan (1996). "Is Fixed Investment the Key to Economic Growth?" *The Quarterly Journal of Economics* 111, 269–276.

Easterly, W., and S. Rebelo (1993). "Fiscal Policy and Economic Growth: An Empirical Investigation." *Journal of Monetary Economics* 32 (2), 417–458.

Ford, R., and P. Poret (1991). "Infrastructure and Private-Sector Productivity." *OECD Economic Studies* 17, 63–89.

Harvey, A.C. (1982). *The Econometric Analysis of Time Series*. Oxford: Philip Allan Publishers.

Hsiao, C. (1981). "Autoregressive Modelling and Money Income Causality Detection." *Journal of Monetary Economics* 7, 85–106.

IBRD: The International Bank for Reconstruction and Development (1994). *World Development Report 1994, Infrastructure for Development*. New York: Oxford University Press.

Lynde, C., and J. Richmond (1992). "The Role of Public Capital in Production." *The Review of Economics and Statistics* Vol. LXXIV, 37–44.

Madden, G., and S.J. Savage (1998). "CEE Telecommunications Investment and Economic Growth." *Information Economics and Policy* 10, 173–195.

Mamatzakis, E. C. (1997). "The Role of Public Sector Infrastructure on Private Sector Productivity in a Long Run Perspective." Working Paper.

Munnell, A. H. (1992). "Infrastructure Investment and Economic Growth." *Journal of Economic Perspectives* 6 (4), 189–198.

Ratner, J. B. (1983). "Government Capital and the Production Function for U.S. Private output." *Economics Letters*, 213–217.

Shah, A. (1992). "Dynamics of Public Infrastructure, Industrial Productivity and Profitability." *The Review of Economics and Statistics* Vol. LXXIV, 28–36.

Tatom, J. A. (1991a). "Public Capital and Private Sector Performance." *Federal Reserve Bank of St. Louis Review*, May/ June 1991, 3–15.

– (1991b). "Should Government Spending on Capital Goods Be Raised?" *Federal Reserve Bank of St. Louis Review*, March/ April 1991, 3–15.

Appendix

Table A.1: Data used in the estimations.

Year	Q_t/k_t	KG_t/K_t	h_t/k_t
1970	0.595	0.410	0.116
1971	0.583	0.405	0.108
1972	0.595	0.404	0.102
1973	0.590	0.399	0.096
1974	0.578	0.394	0.091
1975	0.562	0.388	0.087
1976	0.554	0.387	0.084
1977	0.552	0.390	0.082
1978	0.550	0.398	0.079
1979	0.535	0.405	0.072
1980	0.535	0.410	0.068
1981	0.544	0.413	0.069
1982	0.573	0.416	0.071
1983	0.567	0.418	0.067
1984	0.563	0.420	0.064
1985	0.563	0.422	0.062
1986	0.565	0.422	0.059
1987	0.567	0.423	0.057
1988	0.562	0.422	0.054
1989	0.570	0.405	0.051
1990	0.570	0.435	0.049
1991	0.568	0.442	0.049
1992	0.553	0.457	0.046
1993	0.561	0.475	0.043
1994	0.590	0.492	0.043
1995	0.615	0.500	0.043
1996	0.668	0.508	0.045
1997	0.690	0.514	0.045

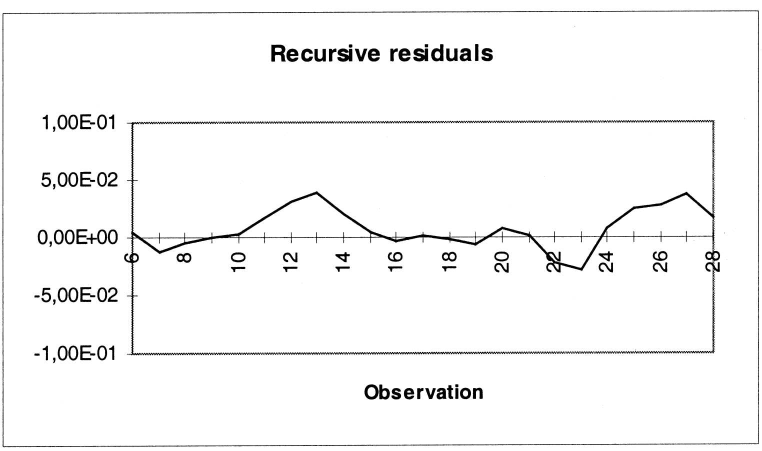

Figure A.1 Recursive residuals from specification (2').

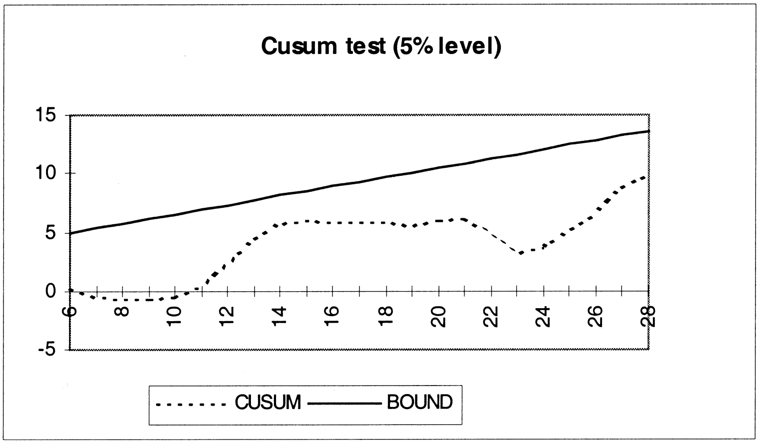

Figure A.2 A cusum test on recursive residuals from specification (2').

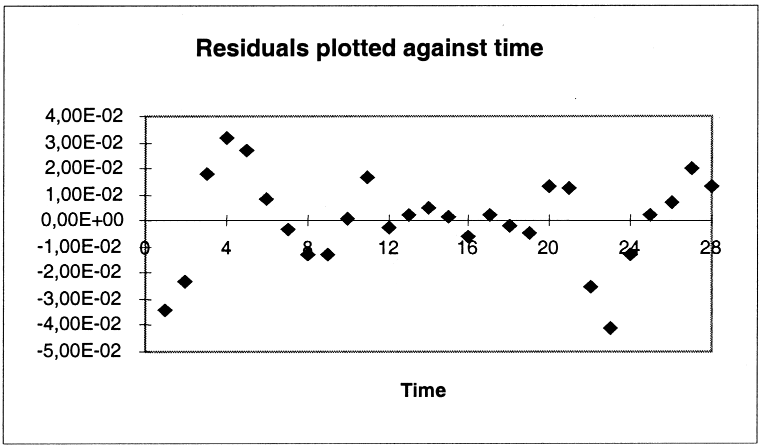

Figure A.3 Residuals from specification (2') against time.

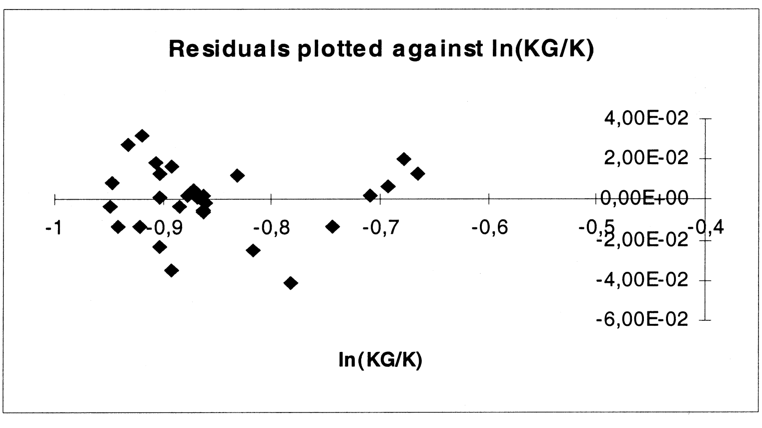

Figure A.4 Residuals from specification (2') against an independent variable.

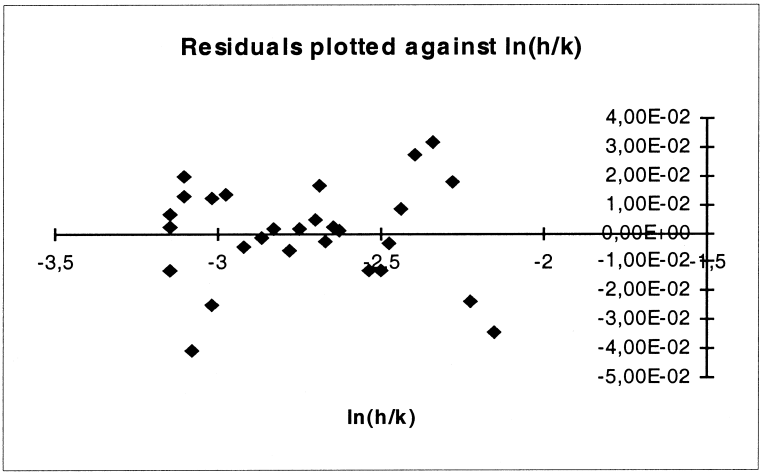

Figure A.5 Residuals from specification (2') against an independent variable.

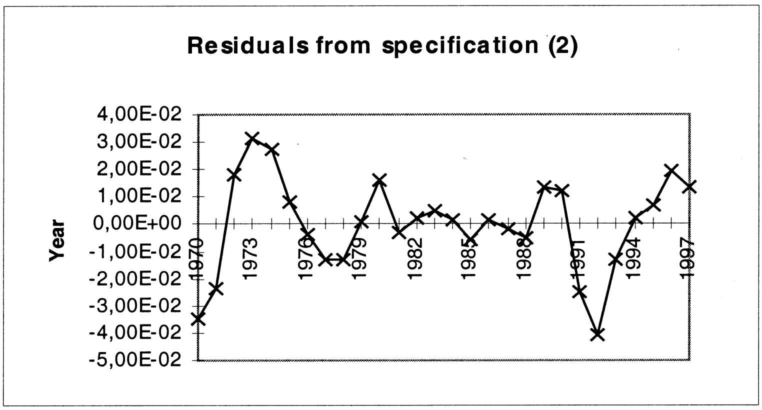

Figure A.6 Serially correlated errors from specification (2).

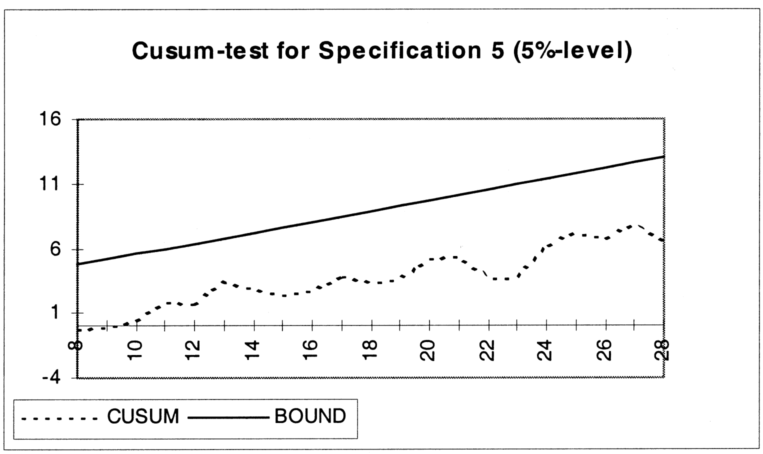

Figure A.7 A cusum test on recursive residuals from Specification (5).

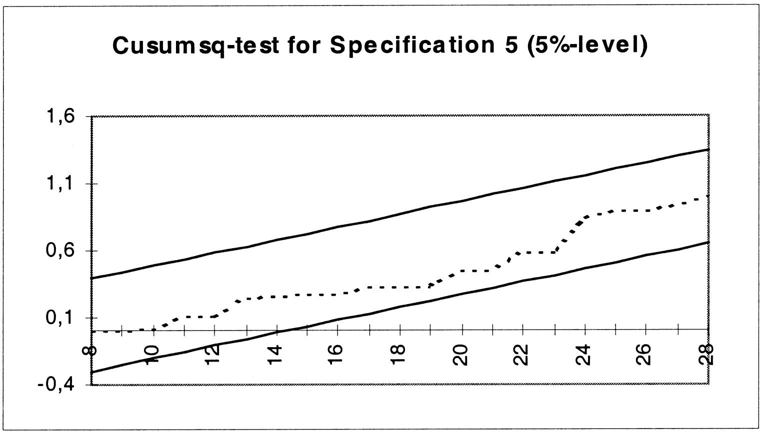

Figure A.8 A cusum sq-test on resursive residuals from Specification (5).