RAND Journal of Economics
Vol. 29, No. 3, Autumn 1998
pp. 596–609

# Monopoly and soft budget constraint

**Ilya R. Segal***

*A benevolent government may decide to subsidize an unprofitable monopoly whose profits do not capture all the social surplus from its production. Anticipating this, the firm may underinvest in order to become unprofitable and extract state subsidies. The resulting welfare loss may exceed by many times the deadweight cost of monopoly pricing. Committing the firm to a price ceiling may soften its budget constraint and thus reduce welfare. Competition can harden budget constraints in industries in which free entry is socially excessive.*

## 1. Introduction

■ By 1993, the Russian economy was in a deep crisis. As prices were liberalized and managers were granted more control over enterprises, many firms found production unprofitable, and the government had to hand out huge subsidies to support them. According to the World Bank (1993), explicit budgetary subsidies to industrial enterprises constituted 22% of GDP in 1992. In the same year, directed industrial credits from the Central Bank of Russia, carrying highly negative real interest rates, cost the equivalent of 21% of GDP. These subsidies were primarily financed by seignorage, which not surprisingly resulted in an inflation of 20% per month.

A large part of the subsidies went to the military-industrial complex, which accounts for a substantial share of Russian industry. When orders from the armed forces collapsed, most factories did not rush to produce for the civilian market, and even when they did, production turned out to be extremely inefficient.[1] In 1992 alone, the Russian government allocated 42 billion rubles ($100 million) for conversion of the military sector. The surprising part, however, is that not a ruble out of these subsidies went to converting machinery or retraining workers (*The Economist,* 1993).

The present article attempts to explain this situation as an extreme manifestation of the soft budget constraint (SBC) syndrome. The concept of SBC was introduced by Kornai (1986), who defined it as a situation in which a loss-making firm can count on a subsidy from the Center. Kornai discussed how SBC may lead to inefficient and unprofitable production. However, although many consequences of SBC have long been

* University of California-Berkeley; segal@econ.berkeley.edu.

This article would not have been possible without many fruitful discussions with Andrei Shleifer. I am also grateful to Mathias Dewatripont, Aaron Edlin, Eric Maskin, Philip Panet, Steven Tadelis, Martin Weitzman, Michael Whinston, and participants of seminars at Harvard and Tel Aviv University for helpful comments. Any remaining errors are my own.

[1] *The Economist* (1993) gives the example of a military factory producing ice-drills for winter fishermen made of titanium (titanium welding is an extremely costly process). Small wonder that most of the military enterprises turned out to be unprofitable.

clear, only recently have its causes come under theoretical investigation. Dewatripont and Maskin (1995), Schaffer (1989), and Schmidt (1996) have developed models in which time-inconsistency of the Center lies at the heart of the SBC syndrome: if the Center were able to credibly commit itself to not subsidize the firm *ex post,* the firm would make more efficient *ex ante* decisions. Although there exist alternative explanations of state subsidies, some of which are discussed in the next section, I believe that in most cases time-inconsistency is an important part of the SBC story.

To tell the full story and draw normative conclusions, one must specify why the Center has an *ex post* incentive to bail out the firm. When the Center is a benevolent government, it wants to correct the *ex post* laissez-faire outcome only if some market imperfection makes this outcome socially inefficient. In the model of Dewatripont and Maskin (1995), firms face an imperfect credit market, which may make a government subsidy *ex post* welfare improving. The present article instead follows Schmidt (1996) in attributing the divergence between the firm's profit-maximizing outcome and the social objective to monopoly power.

A product-market monopoly that is unable to perfectly price discriminate does not capture all the social surplus of production. Alternatively, a firm competing in the output market could enjoy monopsony power in an input market. An important example of such a monopsonized input is labor. Although a detailed description of the labor market is beyond the scope of this article, I can list three reasons for the existence of workers' rents and quasi-rents. First, a firm may not be perfectly informed about individual workers' reservation wages. Second, workers with relationship-specific human capital may have some bargaining power in negotiating wages. Finally, worker rents may be required to induce effort if the harshest punishment for poor performance is dismissal (see, e.g., Shapiro and Stiglitz, 1984). These rents and quasi-rents are not internalized by the firm in the same way that consumer surplus is not internalized by a monopoly. In both cases, social surplus from production exceeds the firm's profits, and if technology exhibits increasing returns, production may be privately unprofitable but socially desirable.

For this reason, a benevolent government may sometimes choose to bail out an unprofitable firm. If the firm in question is a monopoly in the product market, the intervention is justified by the social concern about output, whereas if it has monopsony power over labor, the bailout is justified by the social concern about employment. Although the intervention is socially beneficial *ex post,* the *ex ante* welfare consequences may be disastrous. Anticipating a bailout, the firm may deliberately make its product costly and unwanted by consumers. This would make the firm unprofitable and its threat of shutdown credible. Using this threat, the firm can extract part of the social surplus from the government in the form of state subsidies. Under some conditions, the firm prefers this outcome to independent profitable production. When this happens, social welfare suffers in two ways: through production inefficiency and through the deadweight cost of financing state subsidies. I show that when the elasticity of the demand for monopoly output (alternatively, elasticity of supply of monopsony input) is low, the total welfare loss exceeds by many times the traditional welfare loss from monopoly pricing.

The SBC syndrome may also have important political consequences. I show that a monopolist facing SBC favors price controls and production quotas. Such measures may raise the state subsidy it gets, while reducing the *ex ante* social welfare. Thus, return to a command economy serves the interests of a monopolist facing a SBC, but not the interests of society.

To demonstrate the role of monopoly power in the SBC syndrome, I show that budget constraints are likely to harden when sufficiently many firms compete in the industry. This happens because, as shown by Mankiw and Whinston (1986), free entry

in homogeneous markets is likely to be socially excessive. When sufficiently many firms compete in such a market, the social contribution of the least-efficient firm will be negative. Such a firm would not be able to extract a state subsidy, and it would have to become efficient to earn a positive profit. Therefore, in equilibrium all firms will produce efficiently and will not receive state subsidies.

The rest of the article is organized as follows. Section 2 compares my analysis to existing models of state subsidies. Section 3 develops a model of a monopoly's soft budget constraint. Section 4 investigates the role of competition in hardening budget constraints. Section 5 concludes by discussing some implications of my analysis for economic reforms in former socialist countries.

## 2. Other models of state subsidies

■ Following Dewatripont and Maskin (1995) and Schmidt (1996), I model the SBC syndrome as stemming from time-inconsistency of a benevolent government. I do this to demonstrate that even a benevolent government's intervention may be disastrous for social welfare if the government lacks credibility.

Some studies of state subsidies have rejected the paradigm of benevolent government in favor of a positive theory of government. The most notable study of this kind is Becker (1983), which develops a theory of competition for subsidies among self-interested pressure groups. These groups choose the levels of rent-seeking activity, which determine transfers through an exogenously given ("black-boxed") political influence function.[2] This contrasts with my approach, which explicitly models a particular rent-seeking activity used by one pressure group—a monopolistic or oligopolistic industry—to exert economic pressure on the government (representing the population as a whole).

Becker illustrates his theory with the example of the Chrysler Corporation's bailout by the U.S. government. In his view, the subsidies were caused by the large relationship-specific human capital of Chrysler employees, as evidenced by their high wages. According to Becker, this sunk investment reduces the short-run elasticity of supply and lowers the deadweight cost of subsidies, thereby raising the equilibrium subsidy.

My analysis provides an alternative explanation for the softening of Chrysler's budget constraint. I think that the high wages of Chrysler's workers indicate their significant rents (or quasi-rents on the acquired relationship-specific human capital). Due to these rents, unprofitable production was socially desirable, and the U.S. government subsidized Chrysler to prevent imminent liquidation. In other words, Chrysler was subsidized not because the subsidies were socially cheap, but because not subsidizing would lead to a shutdown, the high social cost of which was not internalized by the corporation. Indeed, the threat of Chrysler's bankruptcy was the primary motivation for federal subsidies (see Reich and Donahue, 1985).[3]

## 3. Monopoly

■ **Basic model.** For concreteness, I focus on a product-market monopoly with competitively supplied inputs.[4] The demand for the firm's output is given by a decreasing

[2] A particular model of "malevolent" government has been studied by Shleifer and Vishny (1994). In the model, a politician derives political benefits from excess employment and bribes a firm to increase employment above the socially optimal level.

[3] We leave it for others to speculate whether the possibility of this outcome had been anticipated by Chrysler's management and had adversely affected their effort, which had in itself pushed the company toward bankruptcy.

[4] This model could be translated one-to-one to the case of a firm that has monopsony power over one of its inputs.

continuous inverse demand function $P(q)$. I assume that the firm's technology exhibits the simplest form of increasing returns to scale: production requires paying a fixed cost $F > 0$, and a constant marginal cost $c \geq 0$.

The firm can undertake an action $a$, which I shall refer to as "investment," before producing, with $0 \leq a \leq A$. This investment may be interpreted as either reducing the marginal cost of production by $a$ or increasing consumer valuation for the product by $a$ so that the inverse demand function becomes $P(q) + a$. All that matters for my results are the reduced-form monopoly profit $\Pi(a)$ and the social surplus $S(a)$ from monopoly production, which are the same under both interpretations. For definiteness, I shall stick to the cost-reduction interpretation.

I assume that action $a$ is costless. This assumption greatly simplifies exposition and gives the most stark result: even though cost can be reduced without expense, the monopolist sometimes does not do it. The investment would make production profitable, making the threat of shutdown noncredible and thereby depriving the monopolist of the opportunity to extract social surplus from the government. My conclusions would only be strengthened if investment were costly.

If the firm chooses to produce, its maximal profit is given by

$$\Pi(a) = \max_{q \geq 0} (P(q) - c + a)q - F. \tag{1}$$

Since in general the monopoly's objective function need not be well behaved, I explicitly assume that the above program has a solution:

*Assumption 1.* The maximum monopoly profit $\Pi(a)$ in (1) is achieved at some output $q^m(a)$.

In particular, this assumption guarantees that the monopoly profit is bounded. Given the monopoly output $q^m(a)$, consumer surplus can be computed as $CS(a) = \int_0^{q^m(a)} P(q)\, dq$, and the total social surplus as $S(a) = \Pi(a) + CS(a)$.

It is well known that monopoly profit falls short of the social surplus from production, and that monopoly output, monopoly profit, consumer surplus, and social surplus are all decreasing in the firm's marginal cost. We therefore have the following proposition (for a proof, see, e.g., Tirole (1988), Ch. 1):

*Lemma 1.* (i) $q^m(a)$, $\Pi(a)$, $CS(a)$, and $S(a)$ are all increasing in $a$. (ii) $\Pi(a) < S(a)$ for any $a$.

To make things interesting, let us assume that production is unprofitable if the firm does not invest at all, and it is profitable if the firm invests in full:

*Assumption 2.* $\Pi(0) < 0 < \Pi(A)$.

After the investment is made, the government observes it and has an opportunity to intervene. I make two additional assumptions that greatly simplify exposition (below I argue that relaxing these assumptions would not affect this section's qualitative results):

*Assumption 3.* The government can observe neither the firm's output nor its price.

*Assumption 4.* The firm has all the bargaining power against the government.

Given Assumption 3, the only variable the government may try to affect is whether the firm produces at all. Given Assumption 4, after choosing its investment, the firm can make a take-it-or-leave-it offer to the government specifying the subsidy it wants in order to produce (I assume that short-term contracts conditioning the subsidy on the

firm's production decision are enforceable). To summarize, the game between the firm and the government can be described as follows:

**Game 1.** *Stage 1.* The firm chooses its investment $a$.

*Stage 2.* The firm makes an offer to the government, specifying the subsidy $s$ it wants in order to produce.

*Stage 3.* The government accepts or rejects the offer.

*Stage 4.* If the offer is accepted, the firm receives the subsidy $s$ and produces. Otherwise, the firm chooses whether to produce.

To complete the description of the game, I need to specify the government's payoff. I assume that the shadow social cost of state funds is $\lambda > 0$; that is, because of the deadweight costs of tax collection, \$1 of tax revenue costs taxpayers $\$(1 + \lambda)$. The benevolent government's objective is to maximize total social surplus, which equals $S(a) - \lambda s$ if the firm produces and subsidy $s$ is paid, or zero if the firm does not produce.

Let $a^0$ denote the level of investment at which the firm just breaks even without subsidies, i.e., $\Pi(a^0) = 0$. It is easy to check that the function $\Pi(\cdot)$ is continuous, hence Assumption 2 ensures that such an $a^0$ exists. According to Lemma 1(ii), $S(a^0) > 0$. Now we are ready to solve Game 1.[5]

*Proposition 1.* The subgame-perfect equilibrium outcomes of Game 1 are given by

(i) If $\Pi(A) < S(a^0)/\lambda$, the firm invests $a^0$, receives a subsidy of $S(a^0)/\lambda$, and produces.

(ii) If $\Pi(A) > S(a^0)/\lambda$, the firm invests $A$, receives no subsidy, and produces.

(iii) If $\Pi(A) = S(a^0)/\lambda$, then both (i) and (ii) constitute equilibrium outcomes.

According to Proposition 1, if the maximum subsidy the firm can obtain by underinvesting exceeds its maximum independent profit, it chooses to underinvest. I shall refer to this outcome as the SBC outcome. When SBC arises, all the social gain from monopoly production is dissipated. This dissipation takes place through two channels: inefficiency in production and the deadweight cost of financing state subsidies. Another consequence of SBC is a reduction in the firm's output (see Lemma 1(i) and note that $a^0 < A$).

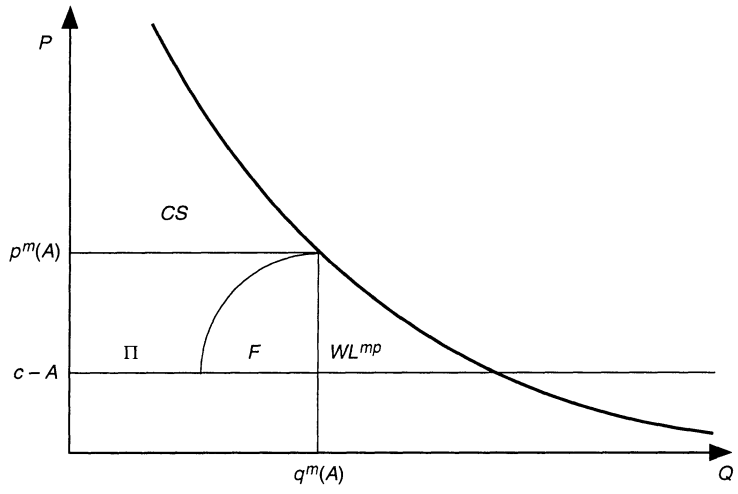
To see how large the welfare loss from SBC (denoted by $WL^{SBC}$) may be, I compare it to the welfare loss from monopoly pricing (denoted by $WL^{MP}$). If the government could credibly commit not to intervene, the firm would invest efficiently ($a = A$), not ask for a subsidy, and sell the monopoly quantity $q^m(A)$ at a price of $p^m = P(q^m(A)) > c - A$. This outcome is depicted in Figure 1.

When the firm has all the bargaining power, SBC reduces social welfare to zero: the social surplus from production is exactly offset by the deadweight cost from financing the state subsidy. Thus, the welfare loss from SBC equals the social surplus from production: $WL^{SBC} = S(A) = CS(A) + \Pi(A)$. Without knowing the fixed cost $F$, we can say little about the monopoly profit $\Pi(A)$, except that it is positive by Assumption 2. For the purposes of estimation, then, let us use the loss of consumer surplus $CS(A)$ as a lower bound for $WL^{SBC}$. Thus, the problem boils down to calculating the ratio $CS(A)/WL^{MP}$. The result obviously depends on the specification of demand. For the standard case of isoelastic demand $Q(P) = P^{-\epsilon}$ with elasticity $\epsilon > 1$,[6] the ratio, $R(\epsilon)$, can readily be calculated, which yields the following proposition:

[5] For proofs of this and all subsequent results, see the Appendix.

[6] When $\epsilon < 1$, monopoly profits are unbounded, i.e., Assumption 1 is violated.

FIGURE 1



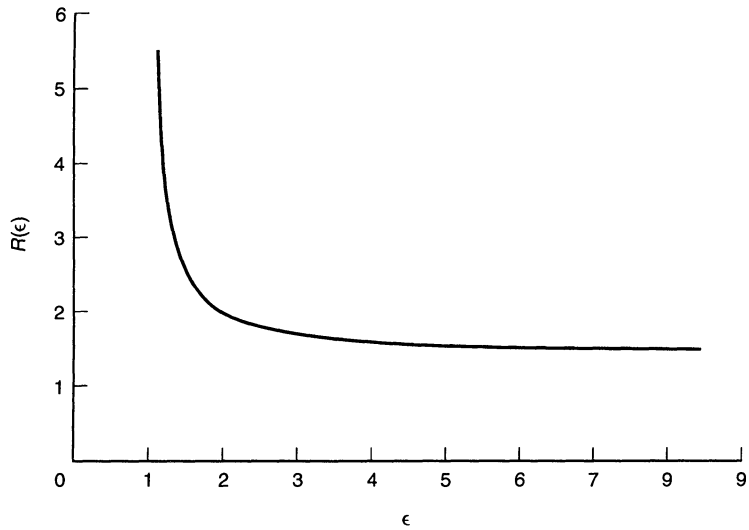

*Proposition 2.* For demand with constant elasticity $\epsilon > 1$,

$$\frac{WL^{SBC}}{WL^{MP}} \geq \frac{CS(A)}{WL^{MP}} = R(\epsilon) = [(1 - \epsilon^{-1})^{1-\epsilon} + \epsilon^{-1} - 2]^{-1}.$$

The graph of $R(\epsilon)$ is presented in Figure 2. It is easy to verify that the function is monotone decreasing, that $R(\epsilon) \to (e - 2)^{-1} \simeq 1.39$ as $\epsilon \to \infty$, and that $R(\epsilon) \to \infty$ as $\epsilon \downarrow 1$. This suggests that the social cost of SBC may be many times higher than the traditional welfare loss associated with monopoly power.

I conclude this section with the observation that none of its qualitative results depend on Assumption 4. Indeed, suppose that stages 2–4 in Game 1 are replaced with

FIGURE 2



an asymmetric Nash bargaining solution, which maximizes a weighted product of the two players' payoffs, $\pi^{\alpha}S^{1-\alpha}$, over their utility possibility set. Here $\alpha \in [0, 1]$ can be interpreted as the firm's bargaining power. In a SBC equilibrium, the firm invests $a^0$, the utility possibility set is given by $\{(\pi, S) | \lambda\pi + S \leq S(a^0)\}$, and the Nash solution yields $\pi = \alpha S(a^0)/\lambda$ and $S = (1 - \alpha)S(a^0)$. SBC arises in equilibrium if and only if $\alpha S(a^0)/\lambda \geq \Pi(A)$. Hence, SBC may arise even when the firm has only partial bargaining power ($\alpha < 1$), even though it is less likely than when the firm has all the bargaining power. Also, observe that when $\alpha < 1$, SBC does not completely dissipate the social surplus from production. However, we can still bound the welfare loss from SBC from below in the case of $\epsilon$-elastic demand. Indeed, using the fact that $CS(a^0) < CS(A)$ (by Lemma 1(i)) and Proposition 2, we obtain

$$WL^{SBC} = S(A) - (1 - \alpha)CS(a^0) \geq CS(A) - (1 - \alpha)CS(A) = \alpha CS(A) = \alpha R(\epsilon)WL^{MP}.$$

Thus, the qualitative results of Propositions 1 and 2 are preserved when the government has some, but not all, bargaining power. However, for the sake of simplicity, in the remainder of the article I restrict attention to the case where the government has no bargaining power against the firm.

□ **Hardening budget constraints.** If the government were able to write a long-term contract with the firm before it invests, conditioning the subsidy on the firm's investment and/or production decisions, this could make the firm invest efficiently. Even if investment were nonverifiable, a simple contract promising the firm a large enough subsidy for producing would keep it from using shutdown as a credible threat and would therefore make it invest efficiently.[7,8] However, following the extensive literature on the dynamics of government regulation, I believe that such long-term contracts are often impractical. Laffont and Tirole (1993) discuss two justifications for this belief. First, most countries have legal prohibitions on the extent to which current administrations can bind future ones. This justification seems especially relevant for politically unstable countries where governments change often. Second (the "incomplete contracting" justification), future technology or production cannot be perfectly described today, which may make long-term contracts meaningless. (For further discussion and examples, see Laffont and Tirole (1993).)

Schaffer (1989) points out that if the government and the firm engage in a long-term relationship, the government can credibly commit itself by developing a reputation for toughness. For the reputation effect to work, however, it is necessary that the government's time horizon be sufficiently long and its discount rate be sufficiently low. The short horizons of governments may explain why the SBC syndrome is most acute in transitional economies.

In this regard, it is interesting to note that after many reformers left the Russian government in a December 1993 compromise with the newly elected parliament, the government became increasingly tough on subsidies and curbed inflation to less than

[7] Furthermore, the government could eliminate subsidies in equilibrium using a contract giving it the right to choose the size of the subsidy $s \geq 0$ *ex post*. The government could then choose $s = -\Pi(a) + \delta$ whenever $\Pi(a) \leq 0$ (with $\delta > 0$ very small), and $s = 0$ otherwise. Thus, the contractual subsidy is indirectly (via the government's choice) contingent on the firm's investment. This "message game" has been suggested by Eric Maskin.

[8] On the other hand, allowing short-term output-contingent contracts would not change the qualitative results of Section 2. With such contracts, even a profitable firm could receive a government subsidy for producing more than the monopoly output. However, an unprofitable firm, which can credibly threaten to shut down, might be able to extract a greater subsidy. Therefore, the firm may still choose to invest $a^0 < A$.

5% per month (*RFE/RL Research Report,* 1994). Since the compromise increased the strength of "industrialists," the interest-group theories of SBC described in Section 2 would have a hard time explaining this phenomenon. On the other hand, a time-inconsistency theory of SBC can attribute it to the increased reputation building by the government whose horizon had been extended by the compromise.

Note also that if the government did not observe the firm's investment, Game 1 would degenerate into a simultaneous-move game, in which it is a dominant strategy for the firm to invest efficiently and the government's best response is not to subsidize. Thus, the government could benefit from a credible commitment not to observe technological information (for example, by dissolving sectoral ministries).[9]

Finally, a government may be able to harden budget constraints by credibly limiting the state budget and committing to a high deadweight cost $\lambda$ of discretionary taxation. (This is most easily seen when $\lambda = \infty$, so the firm cannot expect a state subsidy and invests efficiently.)

□ **Price ceilings/output quotas.** In this subsection I consider a modification of Game 1 in which the firm can commit itself to a price ceiling or output floor observable by the government before the game begins. For example, the firm could write publicly observable long-term contracts with customers, specifying prices and/or quantities (in the labor monopsony interpretation, the firm could write a contract with a trade union committing itself to high wages and/or high employment).[10] This subsection investigates the consequences of such commitments for the firm's profits and for social welfare.

Assuming that demand at a given price is always satisfied, a production quota of $q$ is equivalent to the price ceiling of $P(q)$. For this reason, I focus on the effects of a price ceiling. Normally, a binding price ceiling hurts a monopolist and raises social surplus, and this also happens *ex post* in my model. But when the monopolist has an SBC, an *ex ante* price ceiling may have the opposite effect. To see this, observe that a price ceiling reduces the firm's independent profits and increases the potential state subsidy, which makes SBC more likely. Furthermore, reducing the price ceiling may benefit the firm *ex ante* by increasing the state subsidy it receives. This intuition is formalized in the following proposition:

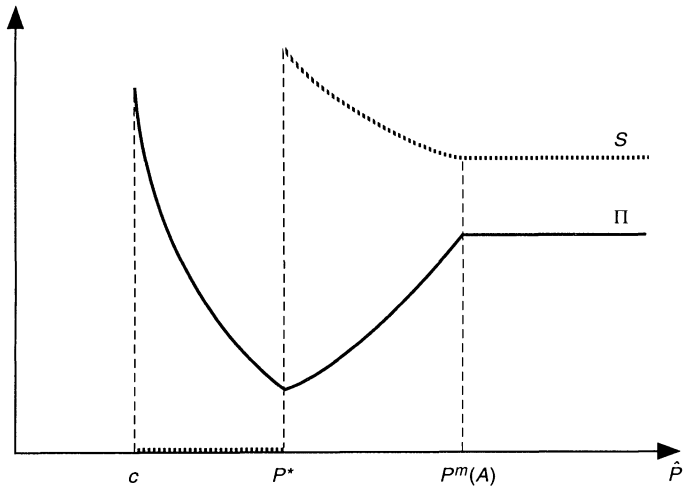
*Proposition 3.* Imposing a ceiling $\hat{p} \geq c$ on the firm's price may soften its budget constraint in Game 1. When the SBC outcome is an equilibrium with an unconstrained firm, imposing a ceiling increases government subsidy and the firm's payoff.

The result is illustrated in Figure 3, drawn under the assumption that without a price ceiling the firm's budget constraint is hard, it invests $A$, and it charges the monopoly price $p^m(A)$. The figure depicts the regulated firm's profits (solid line) and social surplus (dotted line) as functions of the price ceiling $\hat{p}$. When $\hat{p}$ falls slightly below $p^m(A)$, monopoly profits fall and social welfare rises. But when the ceiling falls below some critical price $p^*$, the monopolist chooses to invest $a^0$ and ask for a state subsidy, which reduces social welfare to zero. As $\hat{p}$ decreases beyond $p^*$, the state subsidy rises, and for $\hat{p}$ sufficiently low the subsidy may even exceed the independent monopoly profit. Thus, a firm may want to lobby the government for a price ceiling, counting on state subsidies to make up for the loss of independent profits. A benevolent government, on the other hand, should avoid imposing price ceilings, which could soften firms' budget constraints.

[9] A similar point is made by Schmidt (1996), who believes that this commitment can be achieved by privatization.

[10] Such contracts may be feasible even when long-term contracts with the government are not feasible. As discussed in the previous subsection, governments often face legal constraints on long-term commitment that private parties do not face.

FIGURE 3




The Russian media often portray loss-making producers as complaining that they are unable to "satisfy the demand" for their product without a state subsidy. The simplest way to raise profits and eliminate excess demand at the same time is by raising prices. However, most producers do not favor price decontrols. Instead, they call for government subsidies that would allegedly enable them to expand production and "satisfy the demand." I think that my model sheds some light on this paradoxical situation.

## 4. Competition

■ In this section I investigate under what conditions competition hardens budget constraints. Consider the following game between the government and $N$ *ex ante* identical competing firms:

**Game 2:** *Stage 1.* Each firm $i$ chooses its investment $a_i$ and the subsidy $s_i$ it wants from the government in order to produce.

*Stage 2.* The government decides which offers to accept and which to turn down.

*Stage 3.* Firms whose offers are accepted have to produce, and firms whose offers are turned down decide whether to produce. Producing firms pay the fixed cost $F$.

*Stage 4.* Producing firms compete in the market.

Game 2 is similar to the two-stage entry game analyzed by Mankiw and Whinston (1986). One important difference is that I assume the number of potential producers is limited by the number of existing firms, that is, I consider only the short run where outside entry is impossible.

Entry of an additional firm brings the market price closer to marginal cost, thereby raising social welfare. On the other hand, an entering firm duplicates the fixed cost, which reduces welfare. The first-best solution would involve a monopoly selling at the marginal cost. Given Assumption 3, the government cannot observe either outputs or prices, thus the first-best is unattainable. Therefore, following Mankiw and Whinston (1986), I define the socially optimal number of firms in the industry in the second-best sense, taking the mode of competition as given. Intuition suggests that the government is not going to subsidize production if the number of producers exceeds the socially optimal number. The basic question is whether that many firms can survive in the

industry. If the answer is yes, the government can harden budget constraints by increasing the number of firms in the industry, for example by breaking up existing firms or stimulating entry.

To formalize this intuition, we need some additional notation. Namely, let $N^*(a)$ be the socially optimal number of firms, given that they all have invested $a$. Let $N^0(a)$ be the free-entry equilibrium number of firms in the industry, given that they have invested $a$, i.e., the maximum sustainable number of firms in the industry that guarantees them nonnegative profits.

To solve for the equilibrium we also need to know the outcome of *ex post* competition after the firms have made different investments. Let $S(\mathbf{a})$ stand for the total social surplus when the vector of firms' investments is $\mathbf{a} = (a_1, \ldots, a_N)$, and every firm produces. Firm $i$'s profit in this case is denoted by $\Pi_i(\mathbf{a})$. Also, let $\Delta_i S(\mathbf{a})$ denote the social gain from firm $i$'s production in the case where the firms have invested $\mathbf{a}$. Instead of specifying the mode of competition, I make the following assumption:

*Assumption 5.*

(i) $N^*(a)$ is nondecreasing in $a$.

(ii) $S(\mathbf{a})$ and $\Pi_i(\mathbf{a})$ are anonymous.[11]

(iii) $\partial \Pi_i(\mathbf{a})/\partial a_i > 0$, $\partial \Pi_i(\mathbf{a})/\partial a_j < 0$ for $j \neq i$.

(iv) $\partial S(\mathbf{a})/\partial a_i > 0$, $\partial \Delta_i S(\mathbf{a})/\partial a_j < 0$ for $j \neq i$.

Assumption 5(i) is self-explanatory. Assumption 5(ii) says that "all firms are created equal." Assumption 5(iii) simply says that a firm's investment raises its profit and reduces the competitors' profits. Assumption 5(iv) says that a firm's investment raises social surplus, and that a firm's production is less valuable to society when the competitors have invested more. These assumptions hold for standard models of competition with *ex ante* symmetric firms.

We can now formulate the following proposition:

*Proposition 4.* If $N^*(A) < N \leq N^0(A)$ and Assumption 5 holds, then in every subgame-perfect equilibrium of Game 2, no firm receives a subsidy and all firms (or, possibly, all but one if $\Pi_i(A, \ldots, A) = 0$) invest $A$ and produce.

To understand the logic behind the result, first observe (using Assumption 5(iii)) that when the number of firms does not exceed $N^0(A)$, each firm can guarantee itself positive profits by investing $A$ and producing (for simplicity I focus on the case where $\Pi_i(A, \ldots, A) > 0$). Therefore, in equilibrium all firms must produce, and there is more than $N^*(A)$ of them. But then, using Assumption 5(i, iv), we can see that the least-efficient firm's contribution to social surplus is negative, hence this firm will not be subsidized by the government. Therefore, the least-efficient firm invests $A$, which implies that all firms invest $A$ in equilibrium, become profitable, and receive no subsidies.

The result suggests that whenever $N^* < N^0$ (i.e., free entry is socially excessive), the government can harden budget constraints by choosing $N$, the number of firms in the industry, to satisfy $N^* < N \leq N^0$. Observe that in order to harden budget constraints, it is sufficient to have one firm in excess of what would be socially optimal otherwise, which in most practical cases would be justified by the high social cost of soft budget constraints it prevents. But when is free entry socially excessive? Mankiw and Whinston (1986) show that whenever firms produce perfect substitutes and their mode of competition satisfies three quite general assumptions, $N^* < N^0$ as long as the integer

[11] Formally, for any investment vector $\mathbf{a}$ and any permutation $\pi$ of $\{1, \ldots, N\}$, define a new investment vector $\mathbf{a}_\pi$ such that $(\mathbf{a}_\pi)_i = a_{\pi(i)}$ for all $i$. Assumption 5(ii) then requires that $S(\mathbf{a}_\pi) = S(\mathbf{a})$ and $\Pi_j(\mathbf{a}_\pi) = \Pi_{\pi(j)}(\mathbf{a})$ for all $\mathbf{a}$, $\pi$, $j$.

constraint on the number of firms can be ignored. When they take the integer constraint into account, they show that $N^* \leq N^0 + 1$, i.e., free entry may be socially suboptimal by at most one firm. This covers the important case of a monopoly that may find it unprofitable to enter even though entry is socially desirable.

The integer constraint is unlikely to be important when $N^*$ is sufficiently large. For Cournot competition with isoelastic demand, numerical simulations by Perry (1984) show that entry is socially excessive as long as more than two firms can survive in the industry. For the more tractable case of Cournot competition with linear demand, it can be shown that $N^0 > N^*$ as long as $N^* \geq 2$ (see, e.g., Examples 12.E.1 and 12.E.3 in Mas-Colell, Whinston, and Green (1995)). It thus seems safe to conclude that with the exception of natural monopolies, budget constraints in homogeneous markets can be hardened by competition. This is not true, however, for differentiated markets. From the examples of both spatial differentiation (Salop, 1979) and monopolistic competition (von Weizsäcker, 1980), it is well known that when products are sufficiently differentiated, free entry is socially suboptimal. Intuitively, as differentiation becomes more important, each firm is more like a monopoly in its product niche, and it captures less of the social surplus from its production. In such differentiated markets, competition may not harden budget constraints.

Differentiation seems to be of most concern when the market in question is a labor market. Indeed, labor mobility, both geographic and occupational, is quite costly. In a typical example of SBC, a major employer in a small town threatens to shut down, and the social concern for employment makes the government subsidize the firm. If the firm faces a perfectly competitive output market, its fixed costs have to be covered by paying labor below its marginal product. Two employers may not be sustainable in the town without state subsidies, because competition for labor would bid up wages, and fixed costs of production would not be covered. Thus, a high cost of labor mobility among regions and occupations may generate soft budget constraints that cannot be hardened by competition.

## 5. Conclusion

■ Many economists analyze the economic situation in the former Soviet Republics and Eastern Europe in terms of an inflation-output (or inflation-employment) tradeoff. My article suggests that this approach may be misleading. The tradeoff does exist *ex post,* when a state subsidy may be necessary to support unprofitable production. From the *ex ante* viewpoint, however, soft budget constraints both induce inflationary state subsidies and reduce output. Thus, instead of looking for the right position on a dubious Phillips curve, the government should look for ways of hardening budget constraints.

Some authors (Schmidt, 1996; Shleifer and Vishny, 1994) argue that public ownership is a major cause of soft budget constraints. My article suggests another cause—high industry concentration. My analysis shows that privatization need not harden budget constraints in concentrated industries, and that the resulting welfare loss provides a stronger case for demonopolization than the traditional concern for competitive pricing.

In order to harden budget constraints, the government should identify the market imperfections that trigger the subsidies. For example, when subsidies are caused by the government's concern about an industry's output, the industry's budget constraints can be hardened if sufficiently many firms producing the same output are sustained. (An alternative solution could be to open the market to foreign competition.) Similarly, if the subsidies are caused by the social concern about employment, budget constraints can be hardened if sufficiently many firms can employ workers of the same profession in the same region. The discussion in Section 4 shows that if subsidies are caused by

employment concerns and labor mobility is costly, hardening budget constraints by deconcentration may be problematic.[12] On the positive side, in this case measures that increase labor mobility may be quite instrumental in hardening budget constraints. Such measures may include labor retraining programs, development of housing markets, and abolition of cumbersome institutions of residence registration.

Even when deconcentration may harden budget constraints in an industry consisting of noncooperating firms, there is a danger of collusion. While traditional antitrust analysis has focused on collusion in market pricing, my analysis suggests that collusion in bargaining with the government may be more profitable for firms with soft budget constraints and more detrimental to social welfare.[13] This conclusion may have important implications for developing antitrust institutions in countries in which state subsidies to industries are pervasive.

## Appendix

■ Proofs of Propositions 1–4 follow.

*Proof of Proposition 1.* Game 1 is easily solvable by backward induction. The firm's threat not to produce in stage 4 is credible only if $\Pi(a) \leq 0$, which according to Lemma 1(ii) is equivalent to $a \leq a^0$. The government will accept in stage 3 only if the threat is credible and the social surplus from production outweighs the deadweight loss of financing the subsidy, i.e., $S(a) \geq \lambda s$. In stage 2, therefore, the firm will optimally ask for a subsidy $s = S(a)/\lambda$. Therefore, if in stage 1 the firm invests $a \leq a^0$ to make its shutdown threat credible, it will receive the payoff of $\Pi(a) + S(a)/\lambda$. As both $\Pi(a)$ and $S(a)$ are increasing functions, in this case it is optimal for the firm to invest $a = a^0$, and the firm gets zero profit and the government subsidy $S(a^0)/\lambda$.

On the other hand, if the firm chooses not to make its threat credible, it invests $a \geq a^0$, the government is not going to accept any offer, and the firm receives $\Pi(a)$. In this case the firm optimally chooses $a = A$, receives no subsidy, and obtains the payoff of $\Pi(A)$.

We see that depending on whether the maximum attainable government subsidy exceeds the firm's maximum independent profit, the firm chooses to invest $a$ or $A$, which corresponds to cases (i) or (ii) respectively. If the firm is indifferent, then both subgame-perfect equilibria are possible, which corresponds to case (iii). *Q.E.D.*

*Proof of Proposition 2.* The profit-maximizing monopoly price $p_m$ solves the program

$$\max_p (p - C)Q(p) - F,$$

where $c - A$ is the firm's marginal cost and $Q(p)$ is the demand for its output. For the case of $\epsilon$-elastic demand $Q(p) = p^{-\epsilon}$, the first-order condition of this maximization problem can be written in the famous form

$$(p_m - C)/p_m = \epsilon^{-1}. \tag{A1}$$

Consumer surplus $CS$ (see Figure 1) can then be expressed as

$$CS = \int_{p^m}^{\infty} Q(p)\, dp = p_m^{1-\epsilon}/(\epsilon - 1). \tag{A2}$$

The welfare loss from monopoly pricing $WL^{MP}$ (see Figure 1) can be obtained as

[12] In the early 1990s, concern about output seemed to be a more important cause of state subsidies in Russia than concern about employment. Indeed, in 1992 unemployment in Russia was only 1%, and all estimates for 1993 do not exceed 5%. (EBRD *Annual Report,* 1993).

[13] This may shed some light on the activities of "trade associations"—voluntary coalitions of enterprises created in Russia in place of old ministries, as documented by Joskow, Schmalensee, and Tsukanova (1994). My analysis is supported by the authors' observation that one of the main functions of these associations was lobbying for state subsidies, explicit or implicit (such as protection from foreign competition).

$$WL^{MP} = \int_{C}^{p^m} (Q(p) - Q(p_m))\, dp = (C^{1-\epsilon} - p_m^{1-\epsilon})/(\epsilon - 1) - (p_m - C)p_m^{-\epsilon}. \quad \text{(A3)}$$

The statement of the proposition can now be verified by expressing $p_m$ from (A1), substituting the result into (A2) and (A3), and dividing one by the other. *Q.E.D.*

*Proof of Proposition 3.* For any investment $a$, social welfare may only rise and independent monopoly profit may only fall when the price ceiling is imposed: $S_c(a) \geq S(a)$, $\Pi_c(a) \leq \Pi(a)$ for any $a$. This implies, in particular, that the break-even investment $a_c^0 \geq a^0$. Using Lemma 1(ii) we see that $S_c(a_c^0) \geq S_c(a^0) \geq S(a^0)$, therefore the maximum subsidy the monopolist can get is $S_c(a_c^0)/\lambda \geq S(a^0)/\lambda$, whereas the maximum independent monopoly profit is $\Pi_c(A) \leq \Pi(A)$. The inequalities are strict when the ceiling is binding. In other words, the maximum subsidy rises and the maximum independent profit falls when the ceiling is imposed. According to Proposition 1, this makes the SBC outcome more likely. When the SBC outcome is an equilibrium with an unconstrained monopolist, imposing a price ceiling increases the government subsidy because $S_c(a_c^0) > S_c(a^0) > S(a^0)$. *Q.E.D.*

*Proof of Proposition 4.* As $N \leq N^0(A)$, we know that $\Pi_i(A, \ldots, A) \geq 0$. Define case (i) by $\Pi_i(A, \ldots, A) > 0$ and case (ii) by $\Pi_i(A, \ldots, A) = 0$.

The proof consists of the following four claims:

*Claim 1.* In case (i), in equilibrium all firms produce. In case (ii), at least $N - 1$ firms produce, and firm $i$ may not produce only if $\mathbf{a}_{-i} = (A, \ldots, A)$.

*Proof.* Otherwise a nonproducing firm $i$ can profitably deviate by investing $a_i = A$ and producing:
In case (i): $a_j \leq A$ for $j \neq i$ and Assumption 5(iii) ensures that $\Pi_i(A, \mathbf{a}_{-i}) \geq \Pi_i(A, \ldots, A) > 0$.
In case (ii): $a_j \leq A$ for all $j \neq i$, $a_k < A$ for some $k \neq i$ and Assumption 5(iii) ensure that $\Pi_i(A, \mathbf{a}_i) > \Pi_i(A, \ldots, A) \geq 0$.

*Claim 2.* In case (ii), if one firm does not produce, then no firm receives a subsidy.

*Proof.* By Claim 1, all the producing firms in this case have invested $A$, which makes them strictly profitable and the threat of shutdown noncredible.

Henceforth, I restrict attention to equilibria where all firms produce.

*Claim 3.* In an equilibrium in which all firms produce, (one of) the highest-cost firms does not receive a subsidy.

*Proof.* Let firm $i$ be the highest-cost firm, i.e., $a_i \leq a_j$ for all $j$. Assumption 5(i) implies that $N > N^*(A) \geq N^*(a_i)$. Therefore, by Assumption 5(iv) we have $\Delta_i S(\mathbf{a}) \leq \Delta_i S(a_i, \ldots, a_i) < 0$, which implies that the government is not willing to pay for firm $i$'s production.

*Claim 4.* In an equilibrium in which all firms produce, the highest-cost firm invests $A$.

*Proof.* Otherwise the highest-cost firm (let it be firm $i$), which by Claim 2 receives no subsidies, could deviate profitably by investing $A$, since $\Pi_i(A, \mathbf{a}_{-i}) > \Pi_i(\mathbf{a})$ according to Assumption 5(iii).

Putting together the four claims, we obtain the statement. *Q.E.D.*

## References


BECKER, G.S. "A Theory of Competition Among Pressure Groups for Political Influence." *Quarterly Journal of Economics,* Vol. 98 (1983), pp. 371–400.

DEWATRIPONT, M. AND MASKIN, E. "Credit and Efficiency in Centralized and Decentralized Economies." *Review of Economic Studies,* Vol. 62 (1995), pp. 541–555.

*The Economist.* "Russia: Swords and Ploughshares." Vol. 326 (1993), p. 52.

EUROPEAN BANK FOR RECONSTRUCTION AND DEVELOPMENT (EBRD). "*Annual Report, 1993.*" London: EBRD, 1993.

JOSKOW, P.L., SCHMALENSEE, R., AND TSUKANOVA, N. "Competition Policy in Russia During and After Privatization." *Brookings Papers on Economic Activity: Microeconomics,* (1994), pp. 301–374.

KORNAI, J. "The Soft Budget Constraint." *Kyklos,* Vol. 39 (1986), pp. 3–30.

LAFFONT, J.-J. AND TIROLE, J. *A Theory of Incentives in Procurement and Regulation.* Cambridge, Mass.: MIT Press, 1993.

MANKIW, N.G. AND WHINSTON, M.D. "Free Entry and Social Inefficiency." *RAND Journal of Economics,* Vol. 17 (1986), pp. 48–58.

MAS-COLELL, A., WHINSTON, M.D., AND GREEN, J. *Microeconomic Theory.* New York: Oxford University Press, 1995.

PERRY, M.K. "Scale Economies, Imperfect Competition, and Public Policy." *Journal of Industrial Organization,* Vol. 32 (1984), pp. 313–333.

REICH, R.B. AND DONAHUE, J.D. *New Deals: The Chrysler Revival and the American System.* New York: Times Books, 1985.

*RFE/RL Research Report: Weekly Analyses From the RFE/RL Research Institute.* New York: Radio Free Europe/Radio Liberty, Inc., 1994.

SALOP, S.C. "Monopolistic Competition with Outside Goods." *Bell Journal of Economics,* Vol. 10 (1979), pp. 141–156.

SCHAFFER, M. "The Credible-Commitment Problem in the Center-Enterprise Relationship." *Journal of Comparative Economics,* Vol. 13 (1989), pp. 359–382.

SCHMIDT, K. "The Costs and Benefits of Privatization: An Incomplete Contracts Approach." *Journal of Law, Economics and Organization,* Vol. 12 (1996), pp. 1–24.

SHAPIRO, C. AND STIGLITZ, J.E. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review,* Vol. 74 (1984), pp. 433–444.

SHLEIFER, A. AND VISHNY, R.W. "Politicians and Firms." *Quarterly Journal of Economics,* Vol. 109 (1994), pp. 995–1025.

TIROLE, J. *The Theory of Industrial Organization.* Cambridge, Mass.: MIT Press, 1988.

VON WEIZSÄCKER, C.C. "A Welfare Analysis of Barriers to Entry." *Bell Journal of Economics,* Vol. 11 (1980), pp. 399–420.

THE WORLD BANK. *Subsidies and Directed Credits to Enterprises in Russia: a Strategy for Reform,* Report no. 11782-RU, 1993.