A Tale of Two Reforms

Wei Li*

I identify the absence of well-functioning product markets in transition economies as a sufficient condition under which big bang reduces output initially, while a Chinese-style reform increases output. Big bang dismantles central planning or centralized organization of production, permitting monopolistic and vertically interdependent enterprises to pursue their own monopoly profits by restricting output and inter-firm trade to the detriment of the economy as a whole. The Chinese reform, by maintaining central planning but allowing enterprises to produce for the emerging product markets after they have fulfilled their output quotas under planning, gives enterprises incentives to expand output beyond planned targets.

1 Introduction

With the fall of Communism, the formerly centrally planned economies in Eastern Europe and the former Soviet Union (EEFSU) one by one implemented economic reform, known as “big bang,” aimed at transforming these economies into market economies. The reform involved price liberalization, stabilization, trade liberalization and privatization (Lipton and Sachs, 1990; Blanchard, Froot and Sachs, 1994). The initial result was a sharp decline in output. Between 1989–94, the cumulative output decline in 26 EEFSU countries is estimated at 40.8% (Fischer, Sahay and Vegh, 1996). China adopted a different approach toward transition to a market economy. Instead of abandoning planning initially, China introduced markets

---

*Fuqua School of Business, Duke University; wei.li@duke.edu

This paper is based on part of my Ph.D. thesis submitted to the University of Michigan. An earlier version of this paper was distributed under the title “A Tale of Two Reforms: Why Is ‘Big Bang’ Implosive in Output and Explosive in Prices, While Controlled Marginal Reform Isn’t?” I am grateful to Roger Gordon and Stephen Salant for guidance, to Daniel Berkowitz and David Scharfstein (Editor) for discussion and comments. Thanks are also due to Stephen Coate, Robert Dernberger, Alan Deardorff, Milton Friedman, Simon Johnson, David Li, Peter Murrell, Thomas Rawski, Bruce Reynolds, Mark Schaffer, Greg Shaffer, Matthew Shapiro, Jan Svejnar, anonymous referees and workshop participants at Resources for the Future, University of Maryland, RAND, Hong Kong University of Science and Technology, Simon Fraser University, University of Pennsylvania, London School of Economics, University of Pittsburgh, Hong Kong University, and conference participants at the 1995 ASSA and the 7th World Congress of the Econometric Society in Tokyo 1996. Financial support from CIBER at Duke University, the Mellon Foundation and Academia Sinica in Beijing are gratefully acknowledged.

Copyright © 1999, RAND
at the margin, parallel to planning. This reform, officially known as the “dual-track pricing system,” allowed enterprises to sell their above-the-plan output at market prices after they fulfilled their planned output quotas (Byrd, 1989; Naughton, 1995; Li, 1997a). This reform brought about a rapid and sustained output expansion in the 1980’s in China.

The divergent experience in EEFSU and China is puzzling. Why is big bang, which dismantles an inefficient economic system with the intent of establishing a more efficient one, associated with a large initial decline in economic activity? Why is the Chinese reform, which preserves the inefficient planning system initially, associated with a marked increase in output? In this paper, I identify the absence of well-functioning product markets in transition economies as a sufficient initial condition under which big bang reduces output initially while the Chinese reform increases output.

My theory has three main components. First, I model the Soviet-type economic system as an inefficient but consistent system in which state monopolization of industry and central planning constitute its complementary components. Under state monopolization, central planning is indispensable for integrating a complex network of monopolistic and interdependent enterprises into what Lenin once called “a single factory,” and for coordinating production in the “factory.” Second, I analyze the impact of big bang on a Soviet-type economy in the short-run timeframe in which neither significant improvement in enterprise productivity nor significant entry into capital-, technology- and information-intensive intermediate product sectors has occurred. I show that the disintegration of the centralized organization of production permits each monopolistic state enterprise to pursue its own monopoly profits by restricting output to the detriment of the economy as a whole. The disintegration may thus result in a general reduction in the state industry’s output. Finally, I show that because the Chinese reform preserves the centralized organization of production but allows the emerging product markets, however imperfect they may be, to organize the production of the above-the-plan output, it leads to a general expansion of output in the state industry.

Under central planning, well-functioning, competitive product markets are nonexistent and economic activities are coordinated by a centralized bureaucracy. Industry is monopolized by state-owned enterprises.\footnote{With central planning, monopolization makes sense if economies of scale are important. It also simplifies planning since the planner only needs to deal with a small number of producers.} Production processes are divided into narrow niches along product and/or geographic
lines, with each niche assigned to a single enterprise in most cases. The state industry thus consists of a web of interdependent and monopolistic enterprises. Given this industrial structure, actions taken by an enterprise that restrict output, can have deleterious external effects. Central planning minimizes the externalities by internalizing inter-enterprise transactions and reducing each enterprise’s objective to fulfilling planned quotas. The Soviet-type economic system is thus a coherent system (Ericson, 1991; Kornai, 1992), despite its inefficiency. Given the Soviet-type institutions, the planner or the leadership of the Communist state is in a unique position to exploit their integrated monopoly power. In Section 2, I show that if the planner’s objective is to mobilize resources for rapid industrialization or self-enrichment, then the planner may systematically distort resource allocation in the economy by restricting consumption.

Big bang is designed to be a comprehensive reform (Lipton and Sachs, 1990). By simultaneously implementing stabilization policies, liberalizing prices, breaking up entry barriers, restructuring existing firms, and liberalizing international trade and investment, it was hoped that complementary market institutions and competitive market conducts would emerge quickly (Gates, Milgrom and Roberts, 1996). However, while big bang may be comprehensive ex ante, it loses its complementarity ex post when the economy responds to some reform measures more slowly than to others. Under big bang, price liberalization and decentralization can be achieved almost overnight. But entry into intermediate or heavy industrial sectors, which generally have high capital, technological, managerial, and informational requirements, is expected to occur slowly (McMillan, 1996). In the short run, intermediate product industries are, therefore, likely to be monopolistic or oligopolistic.

I show in Section 3 that the transition under big bang results in decreased output, real wage, and aggregate profits in the short run. The logic resembles the problem of successive monopolies where a switch from vertical integration to decentralization reduces output and integrated profits due to double marginalization. The situation is, however, more complex
here. The switch from centralized to decentralized decision-making allows any price markup or output reduction by any intermediate producer to be propagated through a web of imperfectly competitive producers, resulting in economy-wide decreases in output, real wage and enterprise profitability.

I then show in Section 4 that the Chinese reform can prevent the initial output decline because it forces each enterprise to produce at least its output quota. Moreover, the reform encourages each state enterprise to expand production by permitting it to produce and sell over-the-quota output at a higher market price. The increase in output under the reform implies a reduction in distortion because the residual demands in the emerging markets, which were untapped prior to the reform, are now being served.

This paper is related to Murphy, Shleifer and Vishny (1992) and Blanchard and Kremer (1997). Murphy et al. (1992) analyze the failed partial reform in the former Soviet Union in which input producers can sell at market prices to private firms but must sell to state firms at artificially low state prices. Assuming competitive input markets, they show that the partial reform leads to a diversion of inputs from state firms to private firms, resulting in coordination failures and reduced output and welfare. Their analysis implies that big bang which liberalizes all prices should not reduce output. Blanchard and Kremer (1997) argue that under big bang state firms that are locked into an upstream-downstream bilateral relationship may optimally not offer prices sufficiently high to their crucial suppliers so as to prevent input diversion. The resulting coordinating failures may lead to an initial decline in total output. In this paper, I show that even when state firms offer high enough prices to suppliers to successfully prevent supply diversion, big bang may still lead to a reduction in output.

This paper differs in focus from other theoretical papers on transition. To highlight the comparison between the Chinese reform and big bang, this paper focuses on the commonality between China and EEFSU. For discussions on the implications of the differences in economic structure on the difference in economic performance in transition, see Sachs and Woo (1994) and Qian and Xu (1993). The focus of this paper is also limited to economic factors. For discussions on the political economy issues, see Dewatripont and Roland (1992), Dewatripont and Roland (1996) and Wei (1997).

2 Soviet-type economy

State monopolization of economic and political power in a Soviet-type economy permits the planner—the Communist party leadership—to pursue an
objective that deviates significantly from the collective welfare of the consumers. Here I take as a primitive that the planners’ objective is orthogonal to consumer welfare.

Consider a simple Soviet-type economy consisting of a representative planner and a representative consumer. The economy produces a single final good \( X_0 \), of which \( D \) is consumed by the consumer and the remainder \( G = X_0 - D \) is taken away by the planner. The planner is assumed to maximize \( G \), which he can invest in the military-industrial complex (or enrich himself). The consumer has a labor endowment \( H \), which she allocates between social production \( L \) and leisure \( H - L \). Let the shadow wage rate be the numeraire and \( p_0 \) the shadow price of the final good. The consumer chooses \( D \) and \( L \) to maximize her utility \( U(D,H - L) \) subject to the budget constraint, \( p_0D + H - L = H \). This yields the consumer demand \( D(p_0) \) and the labor supply \( L(p_0) = p_0D(p_0) \). In order to focus on the economic issues, I assume that \( D(p_0) \) and the revenue function \( p_0D(p_0) \) satisfies the conditions listed in Appendix A.

The price \( p_0 \) should be understood as an implicit price such that \( D(p_0) = D^* \) for a given allocation \( D^* \). In a Soviet-type economy, \( p_0 \) typically exceeds the observed official price \( P_0 \) and chronic shortage is common. The difference, \( p_0 - P_0 \), measures the implicit price of the ration coupon or the time and effort wasted in queuing and stockpiling (Weitzman, 1991; Boycko, 1992; Osband, 1992) or the shortage rent which accrues to the planner (Shleifer and Vishny, 1992). In what follows, I adopt the Shleifer-Vishny assumption that the planner captures the shortage rent \( p_0 - P_0 \) so that the implicit consumer price and the implicit planner price are identical to \( p_0 \).

Labor is the only primary factor of production, while services of physical capital and land (measured in flows) are modeled as intermediate inputs. Other natural resources, such as minerals, are extracted and used naturally as intermediate inputs. The planner has full control over the use of all capital, land, and other natural resources, and hence the allocation of all producer goods. The planner, however, has no direct control over the supply of effective labor, which varies with \( p_0 \). He therefore must trade off between allocating labor and natural resources to produce \( D \), which induces labor supply, and allocating these resources to produce \( G \).

The economy uses a roundabout production technology. There are two intermediate product enterprises and one final product enterprise, each pro-

---

4To the extent that the planner cares about consumer welfare in addition to his own consumption, big bang may lead to an even greater fall in output since the output under central planning would be higher. See Li (1997b).

---
ducing a single good. Each intermediate good is produced with a decreasing average cost technology: there is a fixed requirement of $F_i$ units of labor ($i = 1, 2$) and each additional unit of $X_i$ requires $l_i$ units of labor and $a_{ji}$ units of intermediate good $j \neq i$. The final good is produced with a constant-returns-to-scale technology: each unit of $X_0$ requires $l_0$ units of labor, $a_{10}$ units of the first intermediate good and $a_{20}$ units of the second intermediate good. This specification captures the observation that the production of intermediate inputs or heavy industrial products generally has higher capital, technological, managerial, and informational requirements and hence relatively higher overhead costs than the production of consumer products.

To achieve consistency in planning, the planner must balance the production (i.e., the source) and the uses of intermediate goods:

$$X_1 = a_{12}X_2 + a_{10}X_0, \quad X_2 = a_{21}X_1 + a_{20}X_0$$  \hspace{1cm} (1)

Solving for $X_1$ and $X_2$ yields the derived demands for intermediate inputs:

$$X_1 = [(a_{10} + a_{12}a_{20})/(1 - a_{12}a_{21})]X_0 \equiv A_1X_0$$ \hspace{1cm} (2)

$$X_2 = [(a_{20} + a_{21}a_{10})/(1 - a_{12}a_{21})]X_0 \equiv A_2X_0$$ \hspace{1cm} (3)

where $A_i$ is the final good industry’s total unit requirement for intermediate good $i$.\(^5\) “Material” balance in labor requires that the labor supply meets all labor requirements:

$$L(p_0) = l_0X_0 + l_1X_1 + l_2X_2 + F_1 + F_2$$ \hspace{1cm} (4)

Substituting in (2), (3), $X_0 = D + G$, and $L(p_0) = p_0D(p_0)$ yields

$$c_0^*G = (p_0 - c_0^*)D(p_0) - F_1 - F_2$$ \hspace{1cm} (5)

where $c_0^* = l_0 + l_1A_1 + l_2A_2$ is the marginal cost the planner faces in producing the final good.

Equation (5) has an important economic interpretation. The right hand side measures the integrated monopoly profits. Denote it by $S(p_0)$ and rewrite it as $S(p_0) = L(p_0) - c_0^*D(p_0) - F_1 - F_2$. It is thus clear that $S(p_0)$ also measures the labor time over and above that required to produce for the needs of the consumer, or the aggregate labor surplus. The planner’s consumption is therefore financed by the surplus that it extracts from the consumer.

\(^5\)I assume that $1 - a_{12}a_{21} > 0$ so that the planned production is sustainable in the sense that it is capable of generating positive amounts of final output.
Let $p_0(D)$ denote the inverse demand function. Rewriting (5) as

$$G = \left[ (p_0(D) - c^*_0)D - F_1 - F_2 \right] / c^*_0$$

yields the economy’s production possibility locus (PPL), i.e., all feasible combinations of $D$ and $G$ given the technology and the institution. Since labor supply is endogenous, the PPL also represents the amount of available labor at each feasible allocation.

**Centralized equilibrium**

The economy is in a centralized equilibrium when the planner’s action maximizes $G$. Since extracting surplus from the consumer is tantamount to taxing the consumer, the planner’s problem is analogous to an optimal tax problem. Following the convention in the optimal tax literature (e.g., Diamond and Mirrlees, 1971), I restrict the planner to tax only explicit purchases but not the consumption of leisure. The equilibrium is thus second-best.

The first order condition that an interior equilibrium must satisfy is $\partial L / \partial D^* = c^*_0$. That is, the planner equates the marginal increase of labor supply to the marginal cost of eliciting that increase. The first order condition can also be expressed in a Ramsey pricing formula:

$$\frac{p^*_s - c^*_0}{p^*_s} = -\frac{1}{\epsilon},$$

where $\epsilon = D'(p^*_s)p^*_s / D(p^*_0)$ is the elasticity of demand. The markup $p^*_s - c^*_0$ is the implicit tax on consumption. The equilibrium output of each intermediate input is $X^*_i = A_i(G^* + D^*)$ for $i = 1, 2$. Under the assumptions listed in Appendix A, the centralized equilibrium exists and is unique.

**Proposition 1** If the objective of the planner is to mobilize resources for rapid industrialization or for self-enrichment, then it is in the interest of the planner to extract maximum surplus from the consumer by restricting consumption as an integrated monopoly.

This proposition suggests that contrary to conventional beliefs, profits are important in a Soviet-type economy. But profitability must be considered, as Stalin asserted in 1951, “not from the standpoint of individual plants or industries, and not over a period of one year, but from the standpoint of the entire national economy and over a period of, say, ten or fifteen years” (quoted from Ellman, 1989, p. 25). Stalin’s emphasis on “economy-wide profits” highlights the key feature of the Soviet-type economic system—the full integration of economic activities under state ownership and central planning. While the logic of such a system is already outlined in the previous section, it is worth emphasizing here that central planning, in particular, quantity planning is essential to prevent each monopolistic enterprise from

---

6 The production is assumed to be feasible in producing positive amounts of $D$ and $G$; see Appendix A.
maximizing individual profits at the expense of the economy as a whole. For more discussion on the model of centrally planned economy, see Li (1996).

3 Transition economy under big bang

This section analyzes the impact of big bang by comparing the equilibrium patterns of resource allocation before and after big bang. The analysis here focuses on three salient features of big bang—decentralization, price liberalization, and macroeconomic stabilization. The effects of trade liberalization will be discussed in Section 5.

Consider the following transition economy. 1) State enterprises are autonomous and maximize profits. 2) Prices are liberalized. 3) Stabilization policies are in place. 4) No serious restructuring has occurred in the short run, so the inefficient input requirement coefficients $a_{ij}$, $l_i$ and $F_i$ continue to describe the production technology. 5) Each intermediate industry is a monopoly and the consumer industry is perfectly competitive.

Since a formal tax system has yet to emerge, the government continues to rely on enterprise profits to finance its operations. In order to isolate the short-run impact of big bang, I consider a simple case where the government imposes a lump sum tax on each enterprise. To simplify further, I assume that the lump sum tax is set just below the enterprise’s monopoly profits, leaving each enterprise with an arbitrarily small after-tax profits. The government’s tax revenue is thus approximately $\Pi$. The government then spends its tax income by purchasing the final good, $G$, at the market price, $p_0$. To the extent that stabilization policies are effectively enforced, the government faces a hard budget constraint, or $p_0G = \Pi$.

In this decentralized economy, actions taken by the consumer and individual enterprises determine the allocation of resources. The behavior of the

---

7 Alternatively, one may model the enterprise as a labor-managed firm. Since a labor-managed firm can always be approximated by a profit-maximizing firm facing an above-market wage rate in a static model, the current model can be easily modified and its results reinterpreted.

8 This is a simplifying assumption. What matters here is that the intermediate industries are less competitive than the consumer good industry.

9 This lump sum tax does not affect enterprise behavior as long as enterprises stay open in the short run. That governments must rely on distortionary taxes to raise revenue should strengthen this paper’s results, since distortionary taxes in general reduce monopoly output.

10 Allowing the consumer to share some of the profits would complicate the analysis by introducing an additional demand externality through consumer demand, but would not in general alter the main result of the paper that big bang reduces output.
consumer described in Section 2 remains valid except that \( p_0 \) should now be interpreted as the market price of the final good. The behavior of the decentralized enterprises is modeled below as a two stage price game: the intermediate enterprises simultaneously and independently name intermediate prices \( p_1 \) and \( p_2 \), and then perfect competition in the final good market forces \( p_0 \) to equal marginal cost. Each enterprise produces output to meet its demand at the equilibrium prices.

**Decentralized production**

After \( p_1 \) and \( p_2 \) are announced, the marginal costs of producing the three goods are:

\[
c_0(p_1, p_2) = l_0 + a_{10}p_1 + a_{20}p_2, \quad c_1(p_2) = l_1 + a_{21}p_2, \quad c_2(p_1) = l_2 + a_{12}p_1 \tag{6}
\]

Perfect competition in the final good industry leads to marginal cost pricing \( p_0^* = c_0(p_1, p_2) \) and zero profit \( \pi_0 = 0 \). The intermediate monopolies can potentially earn positive profits \( \pi_i = (p_i - c_i)X_i - F_i \) for \( i = 1, 2 \). Equilibrium in each intermediate market requires that the material balance equations (2) and (3) hold. Using these equations, the aggregate profits are

\[
\Pi \equiv \pi_0 + \pi_1 + \pi_2 = (p_0^* - c_0^*)X_0 - F_1 - F_2 \tag{7}
\]

Substituting in (7) into the government budget constraint \( p_0^*G = \Pi \), one finds that the government demand, \( G = S(D)/c_0^* \), is identical to the planner’s demand.\(^{11}\) The big-bang equilibrium allocation, however, differs from the planner’s allocation since decisions are now made by decentralized industries. The total demand for the final good is then:

\[
X_0 \equiv D + G = p_0^*D(p_0^*)/c_0^* - (F_1 + F_2)/c_0^* \tag{8}
\]

The derived demands for intermediate goods are \( X_i(p_1, p_2) \equiv X_i(p_0^*) = A_iX_0(p_0^*) \) \((i = 1, 2)\). Note that the demand function \( X_0(p_0) \) resembles a profit function with zero marginal cost. It is thus possible that \( X_0(p_0) \) is upward sloping when \( p_0 \) is low. This and other properties of the derived demand curves are shown formally in Appendix B.

The economy is in a big-bang Nash equilibrium when competition in the final good industry leads to marginal cost pricing \( p_0^b = c_0(p_1^b, p_2^b) \) and each intermediate industry simultaneously names its own price \( p_i^b \) to maximize \( \pi_i(p_1, p_2) \). Because \( \pi_i(p_1, p_2) = (p_i - c_i)A_iX_0(p_0^*) - F_i \) is continuous in \( (p_1, p_2) \)

\(^{11}\)This is true only when the government taxes away all monopoly profits.
and strictly concave in $p_i$ given the assumptions listed in Appendix A, there exists a pure-strategy Nash equilibrium (Friedman, 1977). An interior pure-strategy equilibrium must satisfy the following first order conditions:

$$\frac{\partial \pi_i}{\partial p_i} \equiv A_i X_0(p^b_0) + (p^b_i - c_i(p^b_j)) A_i X'_0(p^b_0) a_{i0} = 0, \quad i = 1, 2$$

(9)

Will an interior equilibrium ever occur in the non-decreasing portion of the demand curve? The answer is no. If it were that $X'_0(p^b_0) \geq 0$, then each intermediate industry could increase its profits by increasing its price unilaterally. But this is impossible in a Nash equilibrium. The analysis below will thus focus on the downward sloping portion of the demand curve.

There also exists a boundary Nash equilibrium in which the state industry collapses. Consider the following intermediate prices: $\bar{p}_i = (\bar{p}_0 - l_0)/a_{i0}$, where $\bar{p}_0$ is the reservation price for $X_0$. (The existence of $\bar{p}_0$ is shown in Appendix B.) By definition, $\bar{p}_i$ is the smallest $p_i$ that annihilates the final demand regardless of what price the other intermediate enterprise charges. If each intermediate enterprise charges $\bar{p}_i$ ($i = 1, 2$), then no output is produced and no enterprise makes positive profits. But no enterprise can improve its profits by a unilateral deviation. The price vector $(\bar{p}_1, \bar{p}_2, p^c(\bar{p}_1, \bar{p}_2))$ is thus a Nash equilibrium. This boundary equilibrium is, however, Pareto inferior to any interior equilibrium.

The fact that the transition economy is capable of sustaining Pareto-ranked multiple equilibria suggests that coordination failures may be a plausible explanation for the decline in economic activity under big bang. When the economy slips into the boundary equilibrium, an enterprise wanting to produce is not able to do so because it cannot find any supplier or customer with whom to trade. The state industry as a whole ceases production. The boundary equilibrium, however, is not robust. Its existence owes much to the assumption that all inputs must be used in fixed proportions. Li (1997b) shows that if it is possible to substitute labor for intermediate inputs, the boundary equilibrium can be ruled out. In the following discussion, I focus on the implications of interior equilibria where state enterprises continue to operate in the transition economy.

**The economic impact of big bang**

Before comparing big-bang equilibria with the centralized equilibrium, it is instructive to identify two types of pecuniary externalities in the decentralized economy. First, through production linkages, any action taken by an intermediate enterprise that increases price and decreases output will lower
the other intermediate enterprise’s profits by raising its cost and lowering its demand:

$$\frac{\partial \pi_i}{\partial p_j} = -a_{ji}A_iX_0 + (p_i - c_i)A_iX_0'a_{j0} < 0$$

(10)

for \(i \neq j\). The external effects, or circular externalities,\(^{12}\) arise from transactions between enterprises. Second, there are demand spillovers or demand externalities since industrial profits become government demand which in turn contributes to industrial profits. Demand externalities arise from transactions between the government and enterprises.

The effects of the externalities can be seen by differentiating \(G = (\pi_1 + \pi_2)/p_0\) with respect to \(p_j\) \((j = 1, 2)\) and evaluating the result at a big-bang equilibrium:

$$\frac{\partial G(p_0^b)}{\partial p_j^b} \equiv G'(p_0^b)a_{j0} = \frac{1}{p_0^b} \frac{\partial \pi_i}{\partial p_j^b} - \frac{G}{p_0^b}a_{j0} < 0$$

(11)

The first term on the RHS registers circular externalities, while the second term measures demand externalities. These externalities, however, are internalized in the centralized equilibrium, i.e., \(\partial G(p_0^s)/\partial p_j = G'(p_0^s)a_{j0} = 0\) (Section 2). Each individual enterprise maximizing its own profits does not take into account the negative externalities, and therefore tends to make decisions that lead to lower output and lower aggregate profits. The source of the pecuniary externalities is the combination of imperfect competition, inter-industry dependence and decentralized decision-making.

To develop this argument formally, I examine the behavioral implications of the two types of pecuniary externalities separately. Consider first the effects of circular externalities. Because of roundabout production, a price increase by any producer raises other producers’ costs and prompts them to raise their prices, which in turn causes all affected producers to raise prices, leading to a deleterious chain reaction in price-cost escalation. This price-cost escalation curtails inter-enterprise trade, and hence reduces enterprise output and profits. Turn next to the effects of demand externalities. Under central planning, the planner can internalize the production and the consumption of \(G\), and extract surplus at the expense of the consumer only (Proposition 1). Under big bang, each enterprise maximizing its own profits \(\pi_i = (p_i - c_i(p_j))A_i(D + G) - F_i\) extracts monopoly rent at the expenses of both the government and the consumer. Considering that \(G\) is

\(^{12}\)The circular externalities are similar to the vertical externalities in the “chain-of-monopolies” problem (Spengler, 1950).
obtained by levying an implicit distortionary tax on consumption, extending monopoly pricing to $G$ amounts to taxing the consumer twice. The result is an increase in the implicit tax rate or the final price, and a reduction in government spending, consumption and output.

**Proposition 2** If $D(p_0)$ satisfies the conditions listed in Appendix A, then the implementation of big bang leads to the following short-run outcomes:

1. an increase in price or a decrease in real wage ($p^b_0 > p^s_0$),
2. a decrease in output ($X^b_i \equiv X_i(p^b_0) < X^s_i \equiv X_i(p^s_0)$ for all $i$), and
3. a decrease in consumption ($D^b \equiv D(p^b_0) < D^s \equiv D(p^s_0)$),
4. a decrease in government spending ($G^b \equiv G(p^b_0) < G^s \equiv G(p^s_0)$).

**Proof.** Strict concavity of $p_0 D(p_0)$ implies that government spending $G(p_0) = \left[ (p_0 - c^*_0) D(p_0) - F_1 - F_2 \right] / c^*_0$ is strictly concave in $p_0$. $G'(p_0)$ is thus strictly decreasing. At the centralized equilibrium, one has $G'(p^s_0) = 0$. At a big-bang equilibrium, one has $G'(p^b_0) < 0$ (Equation (11)). It thus follows that $p^s_0 < p^b_0$. The remaining claims follow immediately. \( Q.E.D. \)

**Discussion**

The predictions of the model are consistent with the stylized empirical observations in EEFSU regarding the decline in output, the increase in prices relative to wage rate, and the fall in government spending shortly after the implementation of big bang. The model also offers predictions concerning the behavior of state enterprises. In particular, the model predicts that each enterprise will face a contraction in demand and an increase in input prices relative to wage rate. The contraction in demand is attributable to the following factors in this model: the decline in real wage rate, the decline in the government’s real income and the decline in input demand. Surveys of Polish enterprises show that firms perceived a sharp drop in their demand after big bang (Berg and Blanchard, 1994). The increase in input prices relative to wage rate is attributable in this model to monopoly pricing by a “web of monopolies.” The web of inter-industry production linkage propagates and magnifies any intermediate price markup throughout the economy, resulting in higher input prices relative to wage rate. The sharp increase in input costs is indicative of a sharp supply contraction. This prediction is also consistent with empirical observations. In Poland, the initial increase in prices was found to be due primarily to increases in nonwage costs (Berg and Blanchard, 1994).
Transition economy under the Chinese reform

The Chinese reform transforms a Soviet-type economy not to a market economy but to a hybrid economy where plan and market coexist. The state planning authority continues to set a mandatory output quota \( Q_i = X_i(p_s^0) \) for each enterprise and delivers to each enterprise the intermediate inputs \( a_{ji}Q_i \) \((j \neq i)\) required to produce the quota. At the end of each production cycle, the authority procures \( Q_i \) from each enterprise at the artificially low official price \( P_i < p_s^0 \). But planning is no longer the only mechanism through which resources are allocated. The dual-track plan/market system allows state enterprises to produce and sell over-the-quota output, and to buy the required additional inputs at market clearing prices outside of the planning distribution channel. Individual enterprises can now make production and pricing decisions on over-the-quota output. This reform thus creates product markets parallel to planning. With the quotas set at \( Q_i = X_i(p_s^0) \), the planning authority continues to make decisions affecting the intramarginal allocation of resources, but decisions affecting the marginal allocation of resources are decentralized down to individual enterprises.

In this section, I analyze the short-run implications of this reform by focusing on the behavior of the enterprises that are liberalized at the margin, taking as given the planned quotas and planned prices. In order to highlight the comparison between the Chinese reform and big bang, I focus on the commonality between China and EEFSU as transition economies, and study the short-run consequences of applying the Chinese reform to the stylized Soviet-type economy described in Section 2. The transition economy will thus share the same short-run characteristics described in Section 3 with the exception that prices are liberalized only at the margin under the Chinese reform. In order to isolate the short-run impact of the reform, I also maintain the assumption that the government imposes a lump sum tax on each enterprise that takes away (almost) all monopoly profits (see footnote 9).

For the given quotas, actions taken by individual industries determine the allocation of resources outside of planning in conceptually the same way as in a transition economy under big bang. Enterprise behavior will thus be modeled also as a two-stage price game: taking the quotas as given, the intermediate enterprises simultaneously and independently name intermediate prices \( p_1 \) and \( p_2 \) on over-the-quota outputs, and then perfect competition in the final good market forces \( p_0 \) to equal marginal cost. Each industry produces output to meet its demand at the equilibrium prices.
Decentralized production outside of planning

Upon receiving the planned final output \( Q_0 = X_0(p_0^s) \), the government retains \( G^s \equiv G(p_0^s) \) and distributes the rest \( D^s \equiv D(p_0^s) \) to the consumer at the planned price \( P_0 \). Because \( P_0 < p_0 \), the subsidized supply of consumer good must be rationed. After consuming her ration \( D^s \), the consumer can purchase additional units of the final good at the market clearing price \( p_0 \) if her residual demand

\[
d(p_0) = \begin{cases} D(p_0) - D^s & \text{if } p_0 < p_0^s, \\ 0 & \text{if } p_0 \geq p_0^s \end{cases} 
\]

is positive. To meet the consumer’s residual demand, enterprises must produce more than their quotas. In the final good industry, perfect competition in the residual market results in zero residual profit \( \pi_0^r = 0 \) and marginal cost pricing:

\[
p_0 = l_0 + a_{10}p_1 + a_{20}p_2 \quad (13)
\]

For the intermediate monopolies to produce beyond their quotas rationally, however, they must generate additional monopoly profits. The additional monopoly profits, taken away by the government in lump sum taxes, is used to purchase additional units of the final good, denoted by \( g \), for government consumption. The requirement for additional intermediate inputs to produce the residual final output \( x_0 = d + g \) creates the derived residual demand for intermediate goods: \( x_i = A_ix_0 \) \( (i = 1, 2) \). The residual profits that each intermediate enterprise earns by serving the residual demands are:\(^{13}\) \( \pi_i^r = (p_i - c_i)x_i \) \( (i = 1, 2) \). Government’s residual demand \( g \) is thus determined by its budget constraint \( p_0g = \pi_1^r + \pi_2^r \), or

\[
g = (p_0 - c_0^*)d(p_0)/c_0^* 
\]

Similar to (5), this equation represents the trade-off between the government’s residual consumption \( g \) and the consumer’s residual consumption \( d \), or the residual PPL.

Substituting (14) into the residual demand functions yields: \( x_0(p_0) = p_0d(p_0)/c_0^* \) and \( x_i(p_0) = A_ix_0(p_0) \) for \( i = 1, 2 \). Relating the residual demand functions to the unrationed demand functions \( X_i(p_0) \), one can show that

\[
x_i(p_0) = X_i(p_0) - Q_i, \quad i = 0, 1, 2 
\]

\(^{13}\)Note that the fixed costs are already sunk in planned production.
Figure 1: Production Possibility Loci (PPLs) in Soviet-type and Transition Economies.

An enterprise’s residual demand is thus its unrationed demand minus its quota. The residual demand functions $x_i(p_0)$ therefore inherit the properties of $X_i(p_0)$ listed in Appendix B.

Combining (5) and (14) gives the PPL of the dual-track economy for the given quotas:

$$G \equiv G^s + g = G^s + (p_0(D) - c_0^s)(D - D^s)/c_0^s, \text{ for } D \geq D^s \quad (16)$$

Figure 1 depicts the PPL of the Soviet-type economy (and the big-bang economy), and the PPL of the dual-track economy. Point CE, the peak of the Soviet-type PPL, represents the planner’s bliss point under central planning and hence the centralized equilibrium. Point BB, on the upward sloping portion of the Soviet-type PPL represents the big-bang equilibrium. Under the dual-track system, the consumer must supply $L^s = p_0(D^s)D^s$ amount of labor to produce the quota output and is compensated in consumption good $D^s$ in a lump sum. This implies that the PPL starts from point CE. Any incremental labor supply $p_0(D)(D - D^s)$ is compensated by incremental consumption $D - D^s$. The reform thus induces more labor supply, potentially increasing not only consumption but also government spending. Indeed, at point CE, the slope of the dual-track PPL is positive: $(\partial G/\partial D)_{D=D^s} =$
\( p_0(D^s)/c_0^* - 1 > 0 \), whereas the PPL under central planning has a flat slope at point CE. In addition, for all consumption level \( D > D^s \), the dual-track economy is able to generate more \( G \) than the centrally planned economy, since \( G^s + g - [(p_0(D) - c_0^*)D/c_0^* - (F_1 + F_2)/c_0^*] = (p_0(D^s) - p_0(D))D^s/c_0^* > 0 \).

I summarize the result below.

**Proposition 3** The Chinese reform induces an outward shift of the economy’s PPL and potentially moves the economy to a higher level of efficiency.

A decentralized equilibrium in price in the residual markets is a set of prices \( p_{r0}^*, p_{r1}^*, \) and \( p_{r2}^* \) such that

\[
p_{r0}^* = l_0 + a_{10} p_{r1}^* + a_{20} p_{r2}^*
\]

and no intermediate enterprise can increase its profits by unilateral deviation. Concavity and continuous differentiability of \( \pi_i^r(p_1^r, p_2^r) \) \( (i = 1, 2) \) imply that there exists a pure-strategy Nash equilibrium (Friedman, 1977). An interior pure-strategy Nash equilibrium must satisfy the following first order conditions:

\[
X_i(p_{r0}^*) - Q_i + (p_i^r - c_i(p_j^r))X_i'(p_{r0}^*)a_{i0} = 0, \quad j \neq i \quad (17)
\]

There also exists a boundary equilibrium in which each intermediate enterprise charges too high a price to choke off its residual demand, and no enterprise produces beyond its quota. Li (1997b) shows that when the technology is flexible such that additional output can be produced by using additional labor alone, the boundary equilibrium can be ruled out. In the remainder of this section, I focus on the interior equilibria.

**The economic impact of the Chinese reform**

Equation (17) defines for each intermediate enterprise a reaction function, \( p_i^r(p_{rj}^*, Q_i) \). It is apparent that each enterprise’s reaction function is affected directly only by its own quota. By appealing to the second order condition implied by concavity, one can verify the following comparative statics result.

**Proposition 4** An increase in an intermediate enterprise’s output quota shifts its reaction curve downward, and therefore tends to lower its market price and increase its output.

In contrast, however, a marginal increase in quota in the competitive consumer industry, holding all other quotas fixed, will not have any impact on its market price or output.

This proposition implies that the government can influence prices in the imperfectly competitive residual markets by exercising its control over...
output quotas. When every enterprise produces beyond its quota, controlling output quotas amounts to regulating decentralized enterprises *intramarginally*. The intramarginal regulation limits the decentralized decision-making to the residual markets, thereby mitigating the pecuniary externalities that have proven extremely injurious under big bang. To see this, I differentiate \( G = G^s + (\pi_1^r + \pi_2^r)/p_0 \) with respect to \( p_i \) and evaluate the result at the dual-track equilibrium. Doing so yields:

\[
\frac{\partial G(p_0^r)}{\partial p_i} = \frac{1}{p_0^r} \left(-a_{ij}(X_j(p_0^r) - Q_j) + (p_j^r - c_j)X_j(p_0^r)\alpha_{ij}\right) - \frac{a_{i0}}{p_0^r}(G - G^s) < 0
\]

Because the quotas reduce the residual demands, they limit the magnitudes of both circular externalities (the first term) and demand externalities (the second term). But the dual-track system does not eliminate the externalities. Consequently, the equilibrium point CR is on the left of the peak of the Chinese-reform PPL in Figure 1.\(^{14}\)

The equilibrium outcome in a typical intermediate enterprise is depicted in Figure 2. The enterprise must produce its output quota, \( Q_i \). When it produces more than the quota, it is allowed to sell its additional output at

---

\(^{14}\)From (18) one has \( G'(p_0^r) < 0 \). It thus follows that \( G'(D^r) = G'(p_0(D^r))p_0'(D^r) > 0 \), or the dual-track equilibrium is located on the upward sloping portion of the Chinese-reform PPL.
market clearing price $p^*_i$ and earn monopoly profits on the additional sales. Since the additional sales do not lower the planned price on its quota output, the enterprise has an incentive to produce beyond the quota. The Chinese reform, therefore, raises output in every enterprise. The equilibrium market price of the final output $p^*_0$, therefore, must be smaller than the shadow price $p^*_0$ under central planning, implying that the Chinese reform raises real wage. In addition, since each enterprise must earn additional profits when it engages in market transactions by producing output in excess of the quota, the aggregate profits will necessarily increase.

By allowing enterprises to sell their above-the-quota outputs at the market price, the reform in effect permits industries to “price discriminate.” This reform should lead to an efficiency gain because the residual demands, which were untapped prior to the reform, are now being served. In Figure 2, the trapezoidal area ABDE measures the size of the efficiency gain. This gain is shared between the customers and the enterprise or the government. The following proposition summarizes the findings.

**Proposition 5** Given the assumptions listed in Appendix A, if the output quotas are set at the centralized equilibrium levels, then the Chinese reform raises output and profits in every enterprise and raises real wage. The resulting equilibrium Pareto dominates the centralized equilibrium.

These findings can also be understood by relating the Chinese reform to a tax reform. Since government revenues come from monopoly profits both before and after the reform, the monopoly profit margins can be considered commodity tax rates. Under the Chinese reform, the “pre-reform tax rates” or monopoly profit margins apply only to the fixed mandatory quotas, in effect turning pre-reform taxes into lump-sum taxes. If residual markets exist, that is, if $p^*_0 < p^*_0$, then sales in residual markets are taxed at reduced marginal rates $p^*_0 - c^*_0$ (see (16)). The Chinese reform is thus observationally equivalent to a tax reform that reduces marginal tax rates. As a result, enterprises increase production and labor demand; the consumer is offered a higher real wage rate (or more consumption) and hence supplies more labor. The reform, therefore, leads to an increase in aggregate economic activity.

The efficiency gain from the reform is shared between the consumer and the government. This result has two implications. First, the reform has the potential to improve government finance and reduce the fiscal pressure to monetize deficits, thus helping maintain macroeconomic stability during the early stages of transition. The reform experience in China as summarized in McKinnon (1994) is consistent with the model prediction. Second, because the reform delivers an immediate Pareto improvement, it helps build
constituencies favoring more reform, making the reform process sustainable (Dewatripont and Roland, 1992).

Prerequisites for successful implementation

In order for the Chinese reform to bring about a Pareto improvement in the economy, it is necessary that the quotas be strictly enforced. To see why, consider an extreme case where the quotas are not enforced at all. Because market prices exceed planned prices, every enterprise has the incentive to produce only for the market. The resulting equilibrium will be no different from a big-bang equilibrium. To implement the Chinese reform successfully, it is therefore necessary that the government be capable of enforcing the quotas.

The government must also be aware of the rent-seeking opportunities that the divergence of planned and market prices creates for its officials. Corrupt officials may enrich themselves by diverting plan resources under their control to the market. Such official corruption reduces the effectiveness of the dual-track system and undermines political support for the reform. It is therefore also necessary for the government to uphold effective law enforcement in order to limit the damages inflicted by official arbitrageurs. The Chinese government in the 1980s was able to plug enough loopholes in the dual-track system to make it work.

The ability of a government to exert such control is also constrained by the openness of the economy. Before the reform, China was a closed, self-sufficient economy. State control over the economy was nearly absolute. Under these conditions, the Chinese reformers had both the opportunity and the means to implement the reform successfully.

In contrast, EESFU countries were integrated under the CMEA (Council for Mutual Economic Assistance) trade framework prior to the reform, and socialist foreign trade planning was an integral part of central planning. No one country’s central planning authority had complete control over resource allocation within its own border. It is, therefore, doubtful that any one EESFU country could have avoided its own economic decline by unilaterally adopting the Chinese reform. It is also doubtful that the post-communist governments in EESFU would have had the power to enforce quantity quotas after the central planning apparatus had crumbled. However, it remains plausible in theory that the Chinese reform could have succeeded in EESFU if all or most of EESFU countries had coordinated and synchronized their implementations of the reform prior to the demise of central planning.
5 Extensions

A. Input substitution

The fixed technological coefficients in the current model preclude any substitution among inputs. While production is designed to be highly specialized in a Soviet-type economy, forced substitution of inputs has been a fact of life due to chronic shortage and supply disruptions. To capture this substitution flexibility, I extended the model in Li (1997b) to allow for flexible technology. The key results of the exercise are that 1) the extension does not alter the main results of the paper, and 2) the extension shows that the boundary equilibrium discussed in Sections 3 and 4 is not robust and may be ruled out when there is flexibility in input substitution.

B. Wasteful queuing and consumer welfare

Under the assumption that the planner captures all labor surplus (including shortage rents) in the Soviet-type economy, big bang is found to reduce consumer welfare. However, as Boycko (1992) and Osband (1992) have shown, if the consumer must bear some of the efficiency cost of queuing or other rationing schemes by expanding valuable labor time/effort to capture shortage rents, then eliminating shortages in consumer good markets with price liberalization can improve consumer welfare. Under the Chinese reform, this welfare gain should provide additional economic benefit to the consumer. Under big bang, this welfare gain must be compared with the welfare loss arising from imperfect competition in the emerging markets, making it difficult to ascertain theoretically the net welfare change.

C. International trade and the collapse of CMEA

The theory focuses on inter-industry dependence in a closed transition economy. But it can be extended to analyze inter-regionally dependent markets in the presence of imperfect competition, since the inter-industry dependence translates directly into inter-regional dependence if regional economies specialize. The extended theory may offer a plausible explanation as to why trade among the CMEA countries collapsed after big bang. The logic is similar to the paper's explanation of the decline in economic activity under big bang. Notice that the CMEA trade arrangement negotiated by member governments may be economically rational in the eyes of the planners, but not necessarily individually rational to enterprises in each country. As member countries decentralized decisions down to enterprises, the CMEA
trade arrangement could no longer be sustained since individual enterprises would seek to maximize their individual profits, not the planners’ objectives. As a result, not only would there be a reduction in trade among enterprises within each country, there would also be a reduction in trade among enterprises between countries. The extended model should be capable of endogenizing the effects of CMEA collapse on the decline in economic activity among CMEA member countries. In addition, the extended theory would allow one to formally model the impact of trade diversion and import competition on domestic production.

6 Conclusion

Subject to qualifications, the theory developed in this paper offers a consistent explanation of the output decline in EEFSU and the output expansion in China. Its predictions are also consistent with major stylized facts concerning the developments in prices, the aggregate profits and hence government revenue or potential revenue under both reforms.

This paper shows the importance of initial market conditions in transition. Since product markets do not exist under central planning, the short-run performance of a transition economy under big bang depends crucially, among other things, on how fast competitive product markets develop. If the emerging product markets are monopolistic and segmented along production and distribution linkages, big bang is shown to reduce output, lower real wage, and reduce the aggregate profits. The inefficiency in the emerging product markets, however, is not insurmountable. The Chinese approach to transition manages to maintain a delicate balance between plan and market that mitigates the inefficiency in the emerging product markets and results in an expansion of output, an increase in real wage and aggregate profits.

In addition to possibly offering a better understanding of the transition in EEFSU and China, this paper provides a useful lesson for the design of transition policy. If the objective is to have a smooth transition, then a transition policy must be complementary not only among its constituent elements, but also with the economic environment. Complementarity among different elements of the reform itself is necessary but not sufficient.

Appendix A

Assumptions on consumption and production.
1. (Boundedness). $D : p_0 \in [0, \alpha] \rightarrow D(p_0) \in [0, \beta]$, where $\alpha < +\infty$, 
$\beta < +\infty$, $D(0) = \beta$ and $D(\alpha) = 0$;

2. (Differentiability). $D(p_0)$ is twice continuously differentiable in $[0, \alpha]$ 
and $D'(p_0) < 0$;

3. (Concavity). The revenue function $p_0D(p_0)$ is strictly concave in $p_0$.

4. (Feasibility). There exist $p_1$ and $p_2$ such that $D(p_0) = D(l_0 + a_1p_1 + 
a_2p_2) > 0$, $\pi_1(p_1, p_2) > 0$ and $\pi_2(p_1, p_2) > 0$.

Assumptions 1 and 2 implies that there exists a continuously differentiable inverse demand function, $p_0D(p_0)$. Assumption 4 implies that $\Pi(p_0) > 0$ for some $p_0$ so the centralized production is feasible.

**Appendix B**

**Properties of derived demand curves**. Assumptions 1–3 in Appenix A imply that 1) there exists a reservation price $\overline{p}_0 > p_0^*$ for all $X_i$ ($i = 0, 1, 2$), and 2) there exists $\underline{p}_0 < p_0^*$ such that $X_i'(p_0) \leq 0$ for $p_0 \in [\underline{p}_0, \overline{p}_0]$ and $X_i'(p_0) > 0$ for $p_0 \in (0, \underline{p}_0)$ for all $i$.

**Proof**: 1. Let $\mathcal{P} = \{p_0 \in [0, \alpha] : X_0(p_0) \geq 0\}$ denote the support of $X_0(p_0)$. $\mathcal{P}$ is not empty because $X_0(p_0^*) > 0$. $\mathcal{P}$ is bounded from above because $X_0(\alpha) = -F_1 - F_2 < 0$ and $X_0(p_0)$ is continuous. Its least upper bound $\overline{p}_0 = \sup_{p_0} \mathcal{P}$ then defines the reservation price for $X_0$ such that $X_0(\overline{p}_0) = 0$. Because $X_i(p_0) = A_iX_0(p_0)$, $\overline{p}_0$ should also be the reservation price for $X_i$ ($i = 1, 2$). Since $X_0(p_0^*) > 0$, it follows that $\overline{p}_0 > p_0^*$.

2. Strict concavity of $p_0D(p_0)$ (or $X_0(p_0)$) implies that $X_0'(p_0)$ is monotonically decreasing. Using the facts that $X_0'(p_0^*) \equiv G'(p_0^*) + D'(p_0^*) = D'(p_0^*) < 0$ and $X_0(p_0)$ is continuous, one gets $X_0'(p_0) < 0$ for any $p_0 \in [p_0^*, \overline{p}_0]$. Since $X_0'(p_0)$ is monotonically decreasing, only the following two cases are possible for $p_0 \in (0, p_0^*)$:

**Case 1**: $X_0'(p_0) \leq 0$ for all $p_0 \in (0, p_0^*)$. In this case, $\underline{p}_0 = 0$.

**Case 2**: There exists a $p_0^* \in (0, p_0^*)$ such that $X_0'(p_0^*) > 0$. In this case there must exist a $p_0 \in (p_0^*, p_0^*)$ such that $X_0'(p_0) = 0$ and $X_0'(p_0) \leq 0$ for all $p_0 \in [p_0, p_0^*]$.

This proves the second property of $X_0(p_0)$. Because $X_i(p_0) = A_iX_0(p_0)$, the same property holds for $X_i(p_0)$ for $i = 1, 2$.

$Q.E.D.$
References


