

Coercion, compliance, and the collapse of the Soviet command economy

Mark Harrison

Department of Economics

University of Warwick

Coventry CV4 7AL

+44 24 7652 3030 (tel.)

+44 24 7652 3032 (fax)

Mark.Harrison@warwick.ac.uk

Acknowledgements

An earlier draft of this paper was presented to the Economic History Workshop of the University of Warwick, and in a nontechnical version as an inaugural lecture at the University of Warwick, to the Soviet Industrialisation Project Seminar of the University of Birmingham, and to the Centre for Economic History Seminar of Moscow State University. Thanks to Edwin Bacon, Leonid Borodkin, Stephen Broadberry, Nick Crafts, Bob Davies, Michael Ellman, Santayan Ghosal, Gregory Grossman, Philip Hanson, Grigorii Khanin, Vladimir Kontorovich, Byungyeon Kim, Grigorii Khanin, Peter Law, Valery Lazarev, Kalin Nikolov, Andrew Oswald, Geoffrey Renshaw, Gertrude Schroeder, Robert Skidelsky, Stephen White, and others for advice and comments.

This draft: 1 March, 2001. Please do not cite.

Coercion, compliance, and the collapse of the Soviet command economy

Abstract

Are command systems that rest on coercion inherently unstable, and did the Soviet economy collapse for this reason? Postwar evidence is inconsistent with the hypothesis that the Soviet economy was unstable. If it was not unstable, why did it collapse? A repeated game of coordination between a dictator and producers shows that a high level of coercion may yield a stable high-output equilibrium, that the command economy contains a time-consistency problem for central planners, and that a transition to a low state of coercion and performance in which everyone's income falls may be brought about by rising monitoring costs and the dictator's loss of reputation. The facts of the Soviet case are consistent with a collapse triggered when the dictator threw in the towel.

Keywords

command economy

dictatorship

incentives

monitoring

Soviet Union

time-consistency

JEL classification numbers

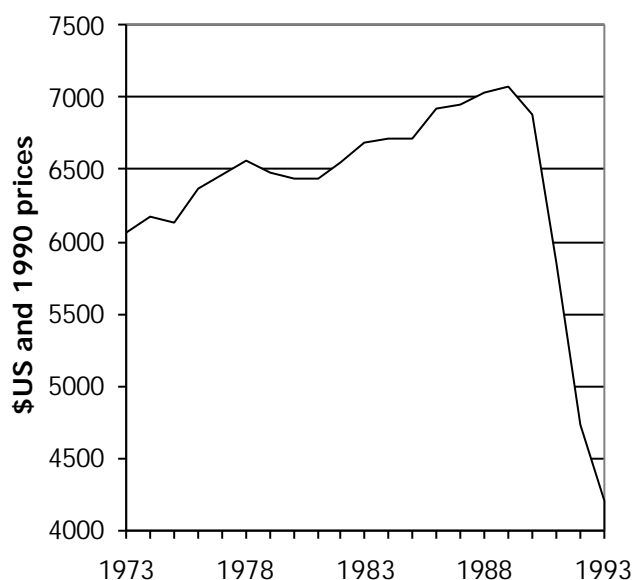
C72, D82, P21

Coercion, compliance, and the end of the Soviet command economy

Introduction

The Soviet economy began to collapse in 1990. As figure 1 suggests, the suddenness with which it did so can scarcely be overstated. After nearly half a century of almost uninterrupted if declining growth, Soviet real incomes fell by one third in three years. After that they fell more slowly, and during the 1990s incomes in Russia remained two fifths below the level of the 1980s.

Figure 1. Soviet real GDP per head, 1973 to 1993



Source: Maddison (1995).

Existing explanations of this collapse are not entirely convincing. A widespread and influential view is that *the Soviet economy collapsed because it was predestined to do so*. 'Essentialists' argue that Soviet society was fundamentally abnormal.¹ Stability requires normality, and normality requires consent, but the Soviet reliance on repression crowded out consent. They maintain that the essence of the Soviet system made its eventual collapse inevitable and predictable; some of them even predicted it significantly beforehand.²

The essentialist argument is a strong one. The Soviet system was repressive, and did collapse. Yet the link from one to the other is usually presented without specifics, and the detail is hard to substantiate. The contemporaneous evidence is that Soviet repression did not crowd out consent; on the contrary there is much evidence of popular support for postwar Soviet institutions. Overt opposition was limited to political dissent and emigration. The dissident movement's narrow social base suggests that in the 1970s most were not interested.³ Surveys of the Brezhnev-era emigration found that émigrés remained loyal to a number of basic Soviet values.⁴ The Gorbachev era provides more substantial evidence of prevailing attitudes and aspirations. Surveys show that most people chose the extent to which they participated in state and party institutions; the more they participated, the more influential they felt over outcomes.⁵ They saw themselves as having more freedoms, with less censorship and less need for self-censorship, than many Americans and most black Americans.⁶ While significant majorities favoured the concepts of *perestroika* and a market economy, most continued to support state ownership of heavy industry and state guarantees of basic incomes and jobs; they did not want consequences of a market economy such as free prices, unemployment, or rich people.⁷

The evidence of the period refutes the idea of an economy waiting to collapse. Time series for 1928 to 1987 show that Soviet productivity was growing.⁸ It rose along a trend which was stable: the economy returned to it when subjected to a disturbance (the latter were frequent and sometimes substantial). The welfare gains realised were large: between 1928 and 1987 GDP per head rose by a factor of five. Real consumption grew by less, and the welfare enhancement permitted by the growing supply of consumer goods and services was diminished by shortages and other restrictions on variety and choice, and by social and intertemporal inequalities. Relative incomes and levels of job satisfaction and general happiness remained low by western standards.⁹ Returns to Soviet accumulation were diminishing, and diminished more sharply than they should have by international standards, especially after the mid-1970s, but they remained positive. Until the mid-1970s the Soviet economy was slowly catching up with the advanced capitalist countries, although still far from overtaking them. Finally, underlying growth fell after the mid-

1970s and became too slow to enable the Soviet economy ever to overtake its rivals, but growth did not fall to zero.

Thus, whether or not the Soviet economy was intrinsically unstable, it grew without any sign of this instability for many years. Indeed one need go back only to World War II to find the Soviet economy displaying much greater resilience under pressure than several more developed market economies.

A popular view is that the postwar Soviet system fell victim to *a growing gap between consumer aspirations and resources*. In particular it is often suggested that the heavy Soviet defence burden worsened this imbalance, and that military–economic rivalry with the United States was contributing to instability. On closer inspection it is hard to take this view seriously. If a wide gap between households' disposable incomes and consumer aspirations was destabilising, few economies would be left standing. And the evidence that the heavy defence burden damaged growth, as distinct from living standards, is not impressive.¹⁰ Indeed it is often overlooked that the Soviet defence industry itself may have been an important source of growth.

Another popular conjecture is that the Soviet economy collapsed because of the spread of *rent-seeking and corruption*. This differs from the essentialist perspective as follows: the Soviet system failed not because of its repressive nature but because repression failed to prevent insider lobbies from dissipating rents and punish investment in lobbying and redistribution rather than production.¹¹ In most accounts rent-seeking and corruption are reasons only for poor performance and slow growth; they do not explain collapse. Exceptions are Murphy, Shleifer, and Vyshny who showed that an economy that starts in a 'good' equilibrium of high output and low rent-seeking may slide to a 'bad' equilibrium in which rent-seekers take over and output collapses. This happens if there is an adverse supply shock and property rights are poorly defended.¹² If so, it is still not clear why an apparently stable regime of the Soviet type with long-term goals and commitments should have ceased to defend state property when the consequences of not doing so were so catastrophic. As will be shown below, an increase in corruption opportunities ought to stiffen the principal's will to monitor agents, not weaken it.

Beforehand many practitioners of Sovietological economics and political science (including the present writer) were sceptical toward prophesies of imminent Soviet economic collapse. After the event they tended to place much weight on *the Gorbachev factor*: the Soviet economy was murdered by caprice, not run down by a deterministic trend.¹³ As Vladimir Kontorovich wrote: 'We tend to confer the mantle of inevitability on accomplished facts, and arguing that what happened did not have to happen is likely to be dismissed as inventing excuses for the losing side. But the collapse of the Soviet system was the unintended result of a small number of disastrous decisions by a few individuals'.¹⁴

Acceptance of a certain role for particular individuals and policies, although perfectly plausible, is consistent with a wide range of opinions about what would have happened without them. At one end of the spectrum Kontorovich asserts that 'were it not for Andropov's kidney disease, communism would still be around'. In the middle, Alexander Dallin also considered that, 'had Gorbachev and his associates *not* come to power, the Soviet Union would have hobbled along, and might have continued to muddle through without overt instability'.¹⁵ Archie Brown concludes only: 'There was nothing inevitable either about *the timing of the end* of the Soviet state or about *the way in which*, under Gorbachev's leadership, the system was transformed' (my emphasis).¹⁶ From another standpoint it may be asked what features of the Soviet system made it so vulnerable to the unintended consequences of the actions of a few and why these consequences, if unintended, were not reversed.

Finally, some explanations for transitional recession have been based on *government policies in central and east European transitions*, but these do not apply to Russia. Jan Winiecki and Janos Kornai have attributed transitional recession to restrictive fiscal and monetary policies.¹⁷ In the Russian case, when real output fell 40 per cent below the previous peak despite persistent budget deficits, monetary growth, and price inflation, the binding constraint was surely supply, not demand. Olivier Blanchard has offered a model of transitional recession involving a supply shock that begins with the removal of subsidies from the state sector.¹⁸ While an adverse effect of removing subsidies is perfectly plausible, it is difficult to apply to the Russian collapse which began in 1990 well before the first subsidies were withdrawn.¹⁹

Why did the Soviet command economy collapse? Are command economies intrinsically unstable? I will argue that stability of a command system is conditional; I will seek to identify the general conditions that may demarcate the command economy's 'good' and 'bad' states, and the particular circumstances that may have pushed Soviet institutions from one to the other. Part 1 considers the nature of economic coercion. Part 2 defines the the players in the command system as self-interested producers and a dictator. Part 3 sets out a coordination game between the players and explores its properties. Part 4 identifies conditions for the command economy's destabilisation. In part 5, the model tells us which stories of Soviet economic collapse are both logical and consistent with known facts, and which are not. Part 6 concludes.

1. Coercion

I take coercion as the central relationship in a command economy. The dictator uses it to prevent the population from working for anyone else; he makes himself a monopolist of capital and a monopsonist of labour. I distinguish economic coercion from repression as follows: economic coercion directs the labour of producers and the choices of consumers and punishes disloyalty to the dictator's economic interests, whereas repression punishes political disloyalty.²⁰ Starting from this, I will develop a model that calibrates coercion, analyses its costs and benefits, and considers the conditions for a high-output solution.

Coercion brings a return that diminishes. Holland Hunter argued that developing economies may gain from a degree of tension in economic planning.²¹ With moderate tension a command economy mobilises resources and grows as a result. As tension rises, returns to coercion diminish because of growing disproportions and errors. Eventually, violence rises to the point where growth declines. Therefore, coercion also has an optimum. In this it is like repression: according to Ronald Wintrobe, political loyalty to a dictator increases in repression up to a point as repression reduces the value of 'disloyal' investments, but beyond that point loyalty declines again because of the rising probability that even behaviour that is intended to be loyal to the dictator will be repressed.²²

Effective coercion rests on the willingness to comply, and securing compliance can be costly. Mills and Rockoff studied the regulation of food supplies in wartime Britain and the USA.²³ They found that compliance was positively associated with the resources invested in coercion. The British enforced food regulations more strictly, detected small infractions more frequently and prosecuted them. As a result there was less free-riding, and food restrictions were widely accepted as equitable. In the United States, monitoring and enforcement were starved of resources and fell below the level that would have secured consent. With rule-breaking endemic and unpunished, the system decayed. Machiavelli would have agreed that, given the will to sustain the costs, coercion can build consent and does not crowd it out. He considered ‘why all armed prophets have conquered, and unarmed prophets have come to grief’. He proposed that ‘the populace is by nature fickle; it is easy to persuade them of something, but difficult to confirm them in that persuasion. Therefore one must urgently arrange matters so that when they no longer believe they can be made to believe by force’.²⁴

In short, coercion brings a return that varies with its intensity. There is a level that is ‘just right’, but coercion may also be too much and too little. Coercion mobilises resources, just as repression mobilises political assets. Its effectiveness relies on the resources invested in monitoring and incentives.

2. The Soviet command system

The dictator

Soviet coercion arose in the context of a command system: production took place in state-owned enterprises regulated by a compulsory plan. Through coercion the dictator became a monopolist of capital and a monopsonist of labour, and was able to mobilise output. He (for dictatorship is invariably a gendered role) returned some of the output to producers for their subsistence on lines recommended by St Paul: ‘he who does not work, neither shall he eat’. The dictator aimed to maximise the surplus or rent that he retained after covering his costs. As a stationary bandit with a long time horizon he invested a considerable part of his surplus in defending and developing the

territory under his control.²⁵ Other costs were the resources he had to invest in monitoring and incentives. By varying the latter the dictator controlled the degree of coercion.

Think of the level of output as dependent on producers' effort in a deterministic way. Producers worked for the dictator, who gave them access to a basic income and additional payoffs conditional upon his monitoring and incentive systems. For this producers had to supply effort, and the supply of effort was what they controlled. The dictator could observe their output if he paid for monitoring. He could not observe producers' effort directly, but while output depended deterministically on effort this did not matter since he could observe output.

In the Soviet command system, who were the producers and who was the dictator? The Soviet command system comprised several multi-level hierarchies with overlapping spheres of responsibility, and this creates significant problems of definition. In what follows the dictator and the producer are defined by their roles: producers controlled effort, while the dictator controlled coercion. The dictator was an economic principal who stood above the law and whose word was law. Planning under dictatorship was not 'law-governed'; rather, 'the plan *is* the law'. The dictator was first and foremost an individual — Stalin, Khrushchev and so on — but some of his powers of decision were delegated to a small group (for the 1930s Paul Gregory calls it 'Team Stalin') of close associates made up by a favoured subset of the party's Politburo.²⁶ In all periods the dictator also required a wider but still numerically small group of economic coordinators and monitors, the central planning staff of Gosplan. So far as has been learnt from the archives at present these acted as the dictator's loyal agents, implementing his decisions and reporting truthfully to him on the outcomes as they saw them.²⁷ In some contexts and periods therefore it will make sense to think of a 'collective dictatorship' comprising a number of individuals bound to each other by common interests.

In the same way there was a hierarchy of producers that extended from the factory workers to managers and officials of fundholding ministries; what they shared was a common interest that lay not in fulfilling the dictator's orders, but in opportunistically influencing the ratio of the rewards the dictator offered for the fulfilment of his orders relative to the effort they supplied to do so.²⁸ At the apex of the system a few of the most important ministers represented the interests

of producers and were also insiders of the dictatorship. The resultant risk of divisions in the Politburo was a permanent preoccupation of Soviet leaders.²⁹ In other words the dictator's agents had an incentive to become roving bandits when the value of an asset they could steal exceeded their shares in the long-run rents from that asset under the dictator.³⁰ The dictator had to counter the incentives for his subjects to rove by forcing them to cooperate. Thus Stalin imprisoned his prime minister Molotov's wife and also invited him to frequent meetings not because he valued his company but to keep him under surveillance.³¹

Monitoring

How was the Soviet command economy monitored? Monitoring made producers account for outputs and inputs and verified its distribution between the dictator's and the producers' uses. Simultaneously, it prevented producers from stealing. Unhindered, producers would steal inputs and products and consume them directly or trade them corruptly, and this would dissipate the dictator's rent. The possibility of stealing rents arose because the dictator had to delegate some control rights over productive assets to the managers and workers who were his agents inside the state-owned enterprises, in the context of a seller's market.

Monitoring was costly, however. The dictator could not check information and enforce decisions without so-called transmission belts: party structures and party-dominated mass organisations at every level of the apparatus and in every workplace. The dictator's planners could not gather products without security guards, transport police, market inspectors, enterprise and ministry accountants, ministry and Gosplan sectors of material balances, Gosbank records, and finance ministry auditors. In fact, planners could not even *count* products particularly well; they had to aggregate them at plan prices, and the definition of real output was subject to inflationary bargaining between producers and planners.³² Thus the dictator had to choose: monitor, and pay monitoring costs; or don't monitor, but let producers steal some output.

Unlike output, monitored imperfectly at some finite cost, effort could be monitored only at a cost that was prohibitive. This follows from evidence of systematic labour-hoarding by enterprises, combined with the fact that official responses to suspected labour-hoarding were not

increased monitoring but revised incentives.³³ The monitoring of effort matters when intrinsic productivities vary and the ratio of output to effort is also randomised. The literature on ratcheting starts from similar premisses.³⁴ However, for present purposes the greater difficulty of monitoring effort is not important: the essential properties of the model are established when effort generates output deterministically.

Incentives

How were incentives designed in the Soviet command system? The dictator offered rewards and punishments. These were ‘artificial’ in that they depended on the dictator’s discretion, not market automatism. The incentives were extremely necessary because, without them, producers would always prefer low effort. Soviet leaders learnt this lesson from the command economy’s formative years 1929–30, when they witnessed a vicious spiral of rising coercion, wage equalisation, and declining productivity.³⁵ After that, the dictator rewarded producers when output was high and punished them when output was low. Rewards were additional income in cash and kind. Punishments were firing and forced labour. Firing, an ineffective threat against most workers under conditions of a general labour shortage, powerfully threatened managers and officials who stood to lose high income, privileges, and careers.

Both threats and promises were costly to the dictator. To meet a promise the dictator had to pay a reward out of his surplus to the producer. To implement a threat involved enforcement costs, which could be very large; at its height, the Gulag employed hundreds of thousands of personnel to detain, transport, and supervise forced labour. To some degree the enforcement costs were covered by the prisoners themselves. For this very purpose labour camps were treated as self-financing cost centres: through forced labour, prisoners and other ‘special settlers’ built their own accommodation and met the costs of their subsistence and detention.³⁶ Indirect punishment costs fell on the dictator, however: both firing and forced labour reallocated workers to jobs of lower intrinsic productivity, and as a result output was lost. The loss varied with the scale of punishments, and was deducted from the dictator’s rents.

Historically command economies have always used both threats and promises. Call a threat or promise that works 'efficient'. At first sight it may not be clear why a dictator should offer promises, if threats would be as efficient. Both were costly if they had to be implemented. To meet a promise meant the dictator had to pay a reward to the producer. To implement a threat required the imposition of a penalty on producers without any corresponding benefit to the dictator, who also had to meet the costs of enforcement. The difference was that efficient promises had to be met (i.e. high output had to be rewarded) whereas efficient threats remained in reserve (when output was high, no one was penalised). If this were all, it would be optimal for the dictator to set penalties at whatever level would make them efficient with rewards set at zero (for example students would always submit essays on time, even for zero credit, if we set the penalty for late submission at death).

In practice the dictator did not have the complete discretion that this required. Suppose there is a maximum penalty that cannot be exceeded. If the maximum penalty for shirking is less than the efficient level, then rewards for not shirking are required as well. An upper limit may be required for penalties to be both feasible and credible. A feasible upper limit is set by the potential income of which producers can be deprived; for this reason positive inducements were significant in labour camps.³⁷ A credible upper limit is set by inherited social norms: if society expects the punishment to fit the crime, unlimited penalties for shirking may not be credible. Specifically, if the ratio of output to effort is partly stochastic and effort cannot be observed so incentives are attached to output, then low effort may be unjustly punished; unlimited punishments may lead to unlimited injustice.

The Stalinist repressions of the late 1930s and 1940s resemble an attempt to combat shirking with unlimited penalties.³⁸ Low effort was classified as 'wrecking' by 'enemies of the people'. Managers and workers suffered harsh penalties for minor production failures. From 1938 and 1940 until after Stalin's death, small labour violations were criminalised regardless of individual circumstances and intentions. Those punished were commonly sentenced to forced labour in establishments subject to self-financing rules under conditions prejudicial to survival. So harsh were the rules that the dictator's normally loyal agents began to refuse to enforce them.³⁹ Stalin's

successors concluded that this regime was inefficient in terms of both incentives and their wider objectives, and shifted the balance from penalties to rewards.

3. A game

Definitions

The following model shows that high coercion can produce a stable state of high output: instability is not intrinsic to a command economy. It reveals a second possible equilibrium characterised by reduced coercion and output — a state of collapse. It defines the conditions under which the economy may slip from a high state to a low state. It shows that as a result everyone may suffer a loss of income.

To avoid a complex parameterisation I assume that the dictator and producers value each unit of current income the same (although they will discount the future differently). The dictator maximises a payoff made up by the value of rents *less* his costs and losses, $Z - M - R - S$, while producers maximise their income received in wages and bonuses and appropriated through theft, *less* the costs of effort and punishments, $Y + S + R - E - P$, defining each symbol as follows:

E	subjective cost of effort	$(E = e, 0)$
F	value of effort, i.e. of the extra output produced by effort	$(F = f, 0)$
M	monitoring costs	$(M = m, 0)$
P	direct cost of punishments	$(P = p, 0)$
Q	productivity cost of punishments	$(Q = q, 0)$
R	value or cost of rewards	$(R = r, 0)$
S	value or cost of stolen rents	$(S = s, 0)$
X	value of output	$(X = x + f, x, x - q)$
Y	value of producer incomes	$(Y = y)$
Z	value of rents	$\left(\begin{array}{l} Z = X - Y \\ = z + f, z, z - q \end{array} \right)$

Levels of variables are set by the dictator and producers as follows. The dictator sets coercion high or low by deciding whether or not to monitor. Without monitoring the dictator cannot stop producers stealing output. The dictator raises coercion by monitoring output, which efficiently eliminates stealing, but monitoring is costly and is a deduction from their rents.

$$\left. \begin{array}{l} M = m > 0 \\ M = 0 \end{array} \right\} \text{and} \left\{ \begin{array}{l} S = 0 \\ S = s > 0 \end{array} \right\} \text{when coercion is} \left\{ \begin{array}{l} \text{high} \\ \text{low} \end{array} \right. \quad (1)$$

When output is monitored, high output can be rewarded and low output punished. Soviet producer incentives were notoriously short-term (Nove, 1958). We will think of them as being applied within the period of the game: high and low output are detected and threats and promises implemented without a lag. When coercion is low, the dictator is indifferent to effort and does not monitor output; coercion is low, not zero, and the dictator still extracts rents, but their level is reduced by producers' theft. Thus,

$$\left. \begin{array}{l} P = 0 \\ P = p > 0 \end{array} \right\} \text{and} \left\{ \begin{array}{l} R = r > 0 \\ R = 0 \end{array} \right\} \text{when coercion is high and output is} \left\{ \begin{array}{l} \text{high} \\ \text{low} \end{array} \right. \quad (2)$$

Output depends on both effort and the scale of punishments. Producers decide whether effort is to be high or low. When effort is low, the value of output is positive and the producer cost of effort is zero. When effort is high and has a positive cost, the value of output is raised by the value of effort. Output depends also on the scale of punishments, because firing and forced labour reallocate workers towards employments of lower intrinsic productivity. Output is high when effort is high, low when effort is low and low output is unpunished, and lower still when low output is punished. (Because planners know who is being punished, the dictator can discriminate between the output loss arising from low effort and that arising as an indirect cost of the punishments he has imposed, so he does not try to punish the latter twice.) For calibration I set $X = x$ when effort is low and unpunished. Thus,

$$\left. \begin{array}{l} E = e > 0 \\ E = 0 \\ E = 0 \end{array} \right\}, \left\{ \begin{array}{l} P = 0 \\ P = 0 \\ P = p > 0 \end{array} \right\}, \text{and} \left\{ \begin{array}{l} X = x + f \\ X = x \\ X = x - q \end{array} \right\} \text{when effort is} \left\{ \begin{array}{l} \text{high} \\ \text{low and coercion is} \\ \text{high} \end{array} \right. \quad (3)$$

Table 1. The dictator controls coercion, producers control effort

		The dictator set coercion	
		High	Low
Producers set effort	High	$z + f - m - r$ Planners' payoff $y - e + r$ Producers' payoff	$z + f - s$ $y - e + s$
	Low	 $z - m - q$ $y - p$	 $z - s$ $y + s$

Finally, the level of rents, the excess of output over wages, is decided as follows; again for calibration I set $Z = z$ when effort is low and unpunished.

$$\left. \begin{array}{l} Z = z + f = x - y + f \\ Z = z = x - y \\ Z = z - q = x - y - q \end{array} \right\} \text{when effort is } \begin{cases} \text{high} \\ \text{low} \end{cases} \text{ and coercion is } \begin{cases} \text{low} \\ \text{high} \end{cases} \quad (4)$$

The payoffs in each case are shown in table 1. Below I will refer to the four corners of the table as follows. The combination of high coercion with high effort is a 'high' state and low coercion combined with low effort is a 'low' state. There are also two mixed states, one with high effort and low coercion, and the other with high coercion and low effort.

Conditions for a high-output equilibrium

A one-shot game

At this point I will show that high coercion can be associated with high output in the command economy in a Nash equilibrium in both the short run and the long run, and the conditions under which this may be the case. In a one-shot game the high state is stable, i.e. none of the parties has an incentive to withdraw from it, provided the dictator does not gain from reducing coercion:

$$m + r < s \quad (5)$$

and at the same time producers do not gain from reducing effort:

$$p + r > e \quad (6)$$

Together these define the feasible space for a command economy: a high–output equilibrium requires the dictator to be able to fix rewards so that:

$$s - m > r > e - p \quad (7)$$

Conversely, if penalties are set too low, or if effort is too costly, producers will shirk regardless of coercion. There is no reward that will motivate them that the dictator is willing to pay. If rents lost to theft are bearable, or if monitoring costs are too high, the dictator will not be bothered to monitor, and he will not wish to offer any reward that producers would accept.

Will either player have a dominant strategy? Consider the situation when condition 5 holds. If producers choose high effort, the dictator will choose to monitor because the return to monitoring, $s - m - r$, must be positive. If producers relax effort rewards are not paid but output falls, so the return to monitoring becomes $s - m - q$. Under condition 5 the latter may be positive or negative depending on whether rewards paid when effort is high are large or small relative to the productivity costs of punishment when effort is low. Thus for the dictator the outcome is ambiguous: when rewards paid are already large and punishment costs are small, high coercion is likely to be his dominant strategy. In the converse case, when producers relax effort the dictator may gain by relaxing coercion, so his strategy will mirror that of producers.

When condition 6 holds, producers will unambiguously follow the dictator, choosing high effort when coercion is high, and low effort when monitoring and incentives are cancelled. Thus a high–output equilibrium is possible only in the presence of high coercion.

A repeated game

When the game is repeated, players consider the discounted stream of future payoffs arising when each other's reaction is taken into account. Define d_D as the dictator's discount factor,

$\left(\frac{1}{1 + \text{discount rate}} \right)$, and d_P as that of producers. Without a capital market producers' discount

factor did not converge with the dictator's. The Soviet dictator's discount factor probably exceeded that of consumers systematically. In the Stalin era the gap was wide enough that the dictator would impose famine on the rural population if necessary to safeguard accumulation

targets, although famine processes were partly attributable to unintended consequences of policy.⁴⁰ After Stalin there were no more famines, and the gap between the dictator's and producers' discount factors perhaps diminished. However, forced saving remains the best explanation of observed consumer behaviours right up to the end.⁴¹

Consider an infinitely repeated game. Start from high coercion and high output. If the dictator reduces coercion and allows producers to steal rents the net immediate change in his payoff will $m + r - s$ immediately, but in a repeated game producers will next respond to low coercion by also cutting effort, giving rise to a further change in the dictator's expected payoff, after discounting, of $\frac{\mathbf{d}_D}{1 - \mathbf{d}_D} \cdot (m + r - s - f)$. The dictator should maintain high coercion as long as the sum of the two would leave him worse off, i.e.

$$s + \mathbf{d}_D \cdot f > m + r \quad (8)$$

This is the long-run condition for high output to be an equilibrium from the point of view of the dictator's willingness to reward producers; $s + \mathbf{d}_D \cdot f - m$ is the maximum reward the dictator is willing to pay, knowing the long-run consequences of abandoning monitoring. This is a higher maximum than in the one-shot game ($s - m$ from condition 5), so the dictator should pay higher rewards to sustain high coercion if he realises or accepts that the game will be repeated.

Now consider the producers' requirement for high effort. If producers shirk, lose rewards, and take punishments, the net immediate change in their payoff is $e - p - r$. The mixed state of high coercion and low effort may or may not be a new equilibrium. The dictator must now meet the indirect productivity cost of punishments. If his costs are too great ($m + q > s$), he may abandon monitoring and move to the low state, so producers' payoff would change again, after discounting, by $\frac{\mathbf{d}_P}{1 - \mathbf{d}_P} \cdot (e + s - r)$. Thus producers should maintain high effort while the sum of the two would leave them worse off, i.e.

$$(1 - \mathbf{d}_P) \cdot p + r > e + \mathbf{d}_P \cdot s \quad (9)$$

This is the long–run condition for high output to be an equilibrium from the point of view of the reward that producers require in return for effort; the minimum reward producers should accept, knowing the long–run consequences of abandoning high effort, is $e - p + \mathbf{d} \cdot (s + p)$. The latter is a higher minimum than in the one–shot game ($e - p$, from condition 6); producers require a higher reward for high effort if they know both the dictator’s costs and that when the game is repeated lower effort may induce the dictator to cut coercion.

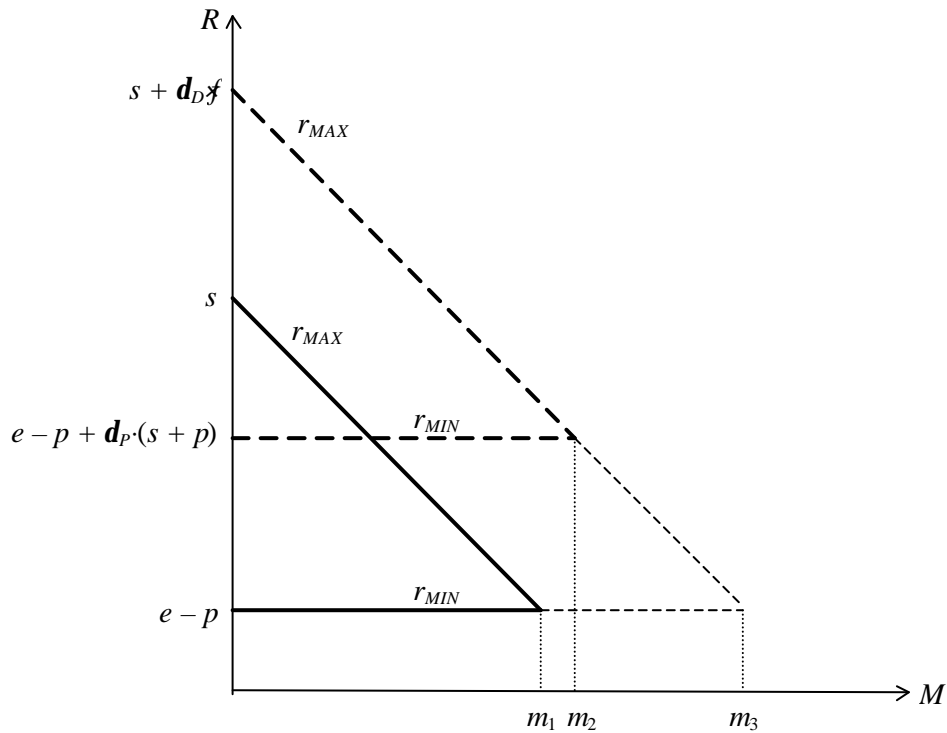
Conditions 8 and 9 yield the feasible space for a command economy in the long run: a high–output equilibrium requires the dictator to be able to fix rewards so that:

$$s - m + \mathbf{d}_D \cdot f > r > e - p + \mathbf{d}_P \cdot (s + p) \quad (10)$$

In the long run the scope for high coercion depends positively on the dictator’s return from high effort, the cost to the dictator of not monitoring, and maximum feasible or credible punishments, and negatively on the costs of effort and monitoring. It also depends positively on the excess of the dictator’s discount factor over that of producers. The more the dictator is orientated towards the long run, the more he will pay to sustain coercion in the present; the more producers are orientated towards the short run, the less they will sacrifice to persuade the dictator to abandon coercion.

Figure 2 illustrates the factors shaping the feasible space for a command economy and reveals the possible triggers for an economic collapse. The one–period space is bounded by the two reward functions making the heavy black triangle. The maximum reward is that payable by the dictator from condition 5, and the minimum reward is that acceptable to producers from condition 6. As long as monitoring costs are less than m_1 , a high–state equilibrium is possible. In the long run, the triangle is shifted upwards to the two reward functions based on conditions 8 and 9, making the heavy dashed triangle. Now the maximum level of monitoring costs consistent with a high–state equilibrium is m_2 . In principle, m_2 could be either greater or less than m_1 , but is shown to be greater in the figure and would be so in fact if $\mathbf{d}_D \cdot f > \mathbf{d}_P \cdot (s + p)$; remember that on the historical evidence \mathbf{d}_D exceeded \mathbf{d}_P .

Figure 2. The feasible space for a command economy



Key: Short run
Long run

Note: $m_3 > m_2$ and $m_3 > m_1$
 $m_2 > m_1$ provided $d_D \cdot f > d_P \cdot (s + p)$

In the long run any combination of the following factors may trigger an economic collapse: a rise in the effort costs of high output or a fall in the value of effort, a reduction in the maximum feasible punishment or an increase in the minimum efficient reward, a rise in monitoring costs or a fall in the scope for producers to steal (i.e. a fall in the costs of not monitoring).

4. Destabilisation

Reputation

Through figure 2 the properties of the game can be further explored. Consider parameter values that permit a solution within the feasible space. As long as rewards are set above the minimum and below the maximum, their value is indeterminate. Under these conditions a process will

emerge in which both sides have something to hide. That their effort is not observed, while the cost of effort is subjective, is to producers' advantage. The dictator must fix rewards by guesswork, not knowing the minimum producers will accept. If he guesses too low, incentives will be inefficient and effort will stay low. If he guesses high, effort will be sustained but he may pay an inefficiently high price for it.

In a repeated game each side can explore the other's limits by trial and error. But the repeated game also introduces extra calculations. In the long run producers may reject a reward that they would have accepted in a one-shot game if they anticipate that, by reducing effort, they may force the dictator to reduce coercion and let them steal rents. The dictator is inhibited by the fear that, if his threat of punishments is ineffective, on the one hand the productivity cost of imposing them may leave him worse off; on the other hand if he does not impose them effort and output will certainly remain low. This fear has the effect of raising the maximum reward he is willing to pay (so the long-run triangle displays an upward shift).

In theory the dictator might have tried to redress the balance by concealing his costs from producers in order to create uncertainty about the maximum reward above which he would withdraw from monitoring and cancel incentives. Thus one could use figure 2 to rationalise Soviet secretiveness in economic affairs: denying information about regime costs to producers could have been one way of keeping costs down. Historically, however, the dictator also adopted a still more obvious way to reduce rewards: that of making a commitment never to relax monitoring under any circumstances. Most famously, in 1937 Stalin declared that mass terror would actually intensify as Soviet society marched nearer to communism; two decades later his successor Khrushchev withdrew this particular commitment but still asserted 'the unshakeable unity of our party, its cohesiveness around the Central Committee, its resolute will to accomplish the great task of building communism [...] We are absolutely certain that our party, armed with the historical resolutions of the 20th congress, will lead the Soviet people along the Leninist path to new successes, to new victories'.⁴² If credible, this commitment would enlarge the feasible space for a command economy to the triangle with a vertex at m_3 by reducing the minimum reward acceptable to producers to that of the one-shot game (condition 8). Producers would

behave as if the game was not repeatable, coercion being set high forever. This might also permit the dictator to persist with a level of monitoring that could turn out to be unprofitable if the game were in fact repeated.

Having given his commitment never to relax coercion, the dictator may gain by behaving time-inconsistently. It may be optimal for the dictator to declare that monitoring will never be abandoned, and low effort will always be punished. Producers who believe this are less likely to relax effort for given incentives. When effort is less likely to be relaxed, at the margin it may become optimal for the dictator to cut his losses by relaxing monitoring and renegeing on potentially costly threats. In other words, when punishments impose a productivity burden there exists a time-consistency problem for central planners in a command economy analogous to that facing central bankers in a market economy: both may trade on their reputation to raise output in the short run.

Since the command economy is not 'law-governed', the dictator cannot bind himself. Consequently the credibility of his commitment never to relax high coercion must rest on reputation alone, and this relies exclusively on an unbroken history. Only if coercion *has* never been relaxed may the belief be sustained that coercion *will* never be relaxed. If coercion is relaxed once, the claim that it can never be relaxed loses credibility forever. Thus reputation is fragile: a single policy error can destroy it.⁴³ This reinforces Machiavelli's dictum that the most dangerous moment for a regime is when it begins to reform itself.

Monitoring costs

Consider monitoring in more detail. The dictator monitored producers to stop them stealing output and dealing on the side. A useful insight is that planners' incentive to monitor was enlarged, not diminished, by corruption opportunities. To the dictator, losing rents was the cost of not monitoring and a reason to make coercion stick. Other things being equal, the bigger the corruption opportunities, the more likely was a high-coercion, high-effort outcome.

But monitoring is costly, and there is reason to suppose that costs were rising. If the difficulty of monitoring grew more rapidly than the monitoring technology improved, then the combined

inputs required to monitor to a given standard rose per unit of final output. Figure 2 shows how a rise in monitoring costs may shrink and finally eliminate the scope for a high-output equilibrium, resulting in a collapse of monitoring and output.

Were monitoring costs rising? Product and process innovations and structural change all pointed in this direction. Growing thousands of heterogeneous products with continuously varied attributes and ceaseless incremental innovation increased planning costs, made output less and less measurable, and impeded the monitoring of productivity. Product innovations assigned from above by powerful user ministries could be monitored directly, but producers channeled much innovation from below into shifting the ratio of reward to effort in their favour in ways that planners could not detect.⁴⁴ As the quality and variety of products grew in economic significance, this strategy proved ever more rewarding. And this is to consider only the 'productive' sphere of the economy. In the growing sphere of non-military services, which raised its share of Soviet employment from one sixth before the war to one third in the 1980s, real output was virtually unmeasured.

Process innovations associated with the transition from mass to flexible production may have fostered producers' control over effort and information. Soviet industry moved from craft production to mass production before and during World War II not only because of its production cost advantages but also to cut monitoring costs.⁴⁵ If mass production began to lose its production cost advantage after the war, flexible production threatened planners with a return to the high monitoring costs previously associated with craft-worker control of production. One could interpret the Soviet postwar failure to engage with flexible production as a decision to forego its growing cost advantages so as to avoid the higher monitoring costs that would come with it.

Is there real evidence of rising monitoring costs? Monitoring costs are hard to identify empirically. For example, from time to time qualified observers expressed the fear that the returns to growth were being eaten up by bureaucracy: for example, in the 1960s a prominent systems analyst predicted that 'at the present rate of development, by 1980 the entire adult population of the USSR will be engaged in planning'.⁴⁶ In fact the proportion of the Soviet population officially

engaged in ‘administration’ remained remarkably constant over many decades at approximately two per cent of public–sector employment.⁴⁷

However, in western economies the indirect regulation costs imposed on those regulated are often substantially more than those paid out of the regulator’s budget. In a Soviet–type economy an important regulator was the communist party itself, which was organised in every workplace and office. Producers were obliged to meet the overheads and direct costs of party facilities and activists’ time. Such outlays, if we knew them, would be a revealing proxy for the trend in monitoring costs. We do not know outlays, only membership. Overall party membership rose steadily in proportion to the working population from less than one per cent in the early 1920s to 3 per cent in 1940, 7 per cent in 1956, 11 per cent in 1973, and 15 per cent in 1986.⁴⁸

Membership was exclusive, based on personal recommendations and a probation period, carried with it an obligation to take part in the work of the party, and was subject to periodic ‘cleansing’.⁴⁹ The rising trend of membership is a clear expression of growing needs.

Economic reforms

One can think of society as consisting of both vertically and horizontally organised networks. A dictatorship is a vertical hierarchy. But production and consumption require transactions that are coordinated horizontally. The classic case of a horizontal network is a market. Hierarchies are weakened by horizontal links because lower level agents can form relationships and collude against superiors.⁵⁰

For transactions involving commodities to take place, information must be exchanged at the same time. One can think of economic development in the twentieth century as imposing a steadily rising information load on transactions. In the 1920s the only information required to order a motor vehicle in the world’s leading market economy was ‘I want a Model T — a black one’. The customisation of products and the growing role of services today require a vastly greater transfer of information for each exchange. It’s cheaper to channel this flow horizontally by the shortest distance from user to supplier than vertically up and down a hierarchy, and this gives horizontal networks their tremendous advantage. For vertical hierarchies a result is that

lower-level agents can form increasingly profitable relationships on the basis of which they will conspire to dissipate the dictator's rents.

Declining Soviet postwar growth with rising costs of both monitoring and *not* monitoring set the scene for a complex cycle of economic reforms in the Soviet Union and eastern Europe.⁵¹ Driving these reforms was the search for a socialist economic mechanism that would regulate itself. It was intended to realign incentives so that plans and producers could coexist with greater harmony than under traditional command arrangements. If reforms succeeded, the dictator could safely delegate control rights to managers without continual costly monitoring.

Improving product allocation was an important goal of economic reform. However, empirical studies of economic reforms suggest that allocation outcomes were often negative. Efficiency was worsened as new incentives provided producers with new avenues for rent-seeking. Hours fell and production discipline was relaxed. Monitoring and controls were then reimposed to correct the consequences. Meanwhile the original problems had not been solved, so a cycle of reforms and counterreforms resulted. Thus reforms failed to stem the rise in monitoring costs.

In the cycle of reforms and counterreforms, shocks towards either too much or too little monitoring prompted reversals. Did the opposing shocks cancel out, with counterreforms enabling the dictator exactly to recapture the power devolved in each reform phase? There is evidence of slippage. Although the market was not strengthened, the plan was gradually weakened. Post-Stalin regimes in the USSR and eastern Europe increasingly tolerated sideline economic activities, which sometimes reallocated resources more efficiently but at the same time undermined state ownership rights; James R. Millar described this as Brezhnev's 'Little Deal' with the Soviet Union's urban population.⁵² In the words of Mancur Olson, state-owned enterprises became 'more nearly insider lobbies or organised special interests than productive enterprises'.⁵³

In short, each reform helped weaken the credibility of the dictator's commitment to coercion. This reputation could not be rebuilt by counterreforms because it relied on an unbroken history. Rising monitoring costs associated with the changing production system, and declining credibility associated with the reform cycle, together squeezed the feasible space for the command economy.

In figure 2 the dictator could profitably sustain monitoring costs up to the level of m_3 on the strength of his reputation alone. As his reputation diminished, the risk increased that he might suddenly find a given level of monitoring costs that was formerly sustainable, say, less than m_3 but more than m_2 , to lie outside his feasible space.

5. Collapse

The first move

When the feasible space for a command economy disappears, the dictator can no longer fix rewards at a level that will induce producers to maintain high effort while generating a positive return to monitoring. At that point output will fall.

The first move decides the sequence. Suppose monitoring costs rise until monitoring is no longer profitable for given rewards. The dictator must reduce rewards or cancel monitoring. In one case *the dictator cuts rewards*. Producers resist by cutting effort, so the economy passes through a temporary punishment phase. Output falls because producers cut effort and because the dictator imposes penalties that cause a further loss of productivity. If the dictator's combined monitoring and productivity costs are large relative to the costs of not monitoring ($m + q > s$), the dictator will then cancel monitoring anyway.

This story is consistent with some facts and not others. It is consistent with evidence of failing rewards. Of nearly 3,000 Brezhnev-era emigrants surveyed by Paul R. Gregory, three quarters reported the impression that average productivity was falling (although it was not); of these, three fifths listed inadequate incentives as the main cause of productivity problems.⁵⁴ It is consistent with the observation that before 1983 work discipline was failing.⁵⁵ But it cannot explain why failing effort was going unpunished. Nor does it explain why, when Soviet coal miners unprecedentedly went on strike for higher rewards under Gorbachev, in the summer of 1989, their withdrawal of effort was not only not punished but rewarded with concessions.⁵⁶ Thus neither period shows the increased penalisation of low effort that is a requisite symptom when failing rewards trigger collapse; on the contrary under Brezhnev plans became less demanding

and penalties declined; under Gorbachev the strikes of 1989 went unpunished.⁵⁷ It is true that in between, in 1983–6, there was a phase of heightened monitoring and discipline. However, this period did not show the falling effort that is a requisite symptom of a collapse triggered by failing rewards; on the contrary effort probably increased, showing that incentives were becoming more efficient. In fact, Soviet output did not begin to fall until 1990; the output lost from failing effort in the years beforehand was not large compared with what was to follow, and in the years when effort was failing there were no productivity costs of punishments since penalties were not being imposed.

In the other case *the dictator abandons monitoring*: rather than cut rewards pointlessly he surrenders first. This at least lets him realise a temporary bonus from high output without paying for monitoring or rewards, before effort falls and the economy collapses.

This story *is* consistent with a much wider range of facts than the previous account. It is consistent with the evidence of failing rewards before 1983 and after 1986. It is consistent with evidence that the stiffening of monitoring and penalties in between rendered rewards temporarily more efficient, so that effort increased, but it suggests that the costs of closer monitoring proved unsustainable. Finally it is consistent with the evidence that Gorbachev's *perestroika* amounted to an unforced dismantling of the institutions used to monitor, reward, and punish producers.⁵⁸ In short, as a former Soviet official told William Keegan: 'We used to work in a centrally controlled system where they told you what to produce. Now they've stopped telling us what to produce, so we don't produce anything'.⁵⁹

The dictator surrenders

There are three possible explanations of the Soviet decision to abandon a command economy.

Each places players' motives and calculations in a slightly different light.

Rational expectations and deterministic trends. Suppose all players know the long-run values of all the variables in condition 10. Adverse trends in monitoring costs and returns to effort shrink the feasible space for a high-output equilibrium to nothing. In the long run monitoring will make a loss, and the dictator rationally closes it down.

Although it did not force the dictator to abandon monitoring, the strike wave of 1989 may have provided him with useful information. How was the dictator to know that monitoring had reached the point of making a loss? By finding that the maximum reward he could offer producers for high effort was no longer efficient. Economic reform having failed to reduce monitoring costs under Brezhnev, planners under Andropov and Chernenko tried to offset rising monitoring costs and failing rewards by increasing penalties. Under Gorbachev, however, increased penalties encountered social and political limits and were eventually abandoned. Lower penalties might lead producers to demand higher rewards, as in the strikes of 1989. If penalties could not be imposed and rewards could not be increased, the strikes served notice on the regime that incentives could no longer be efficient. In summary, the dictator abandoned the command economy, but producers signalled that the time had come. The process of ‘power conversion’ through private capitalisation of party and komsomol networks and enterprises passed the point of no return.⁶⁰

Biased expectations and policy mistakes. If low effort and low coercion are infrequently encountered, players may acquire biased expectations of the values of the variables in condition 10. For example, producers may overestimate the value of rents that they might steal if they reduced effort and induced the dictator to abandon monitoring; this causes them to demand higher rewards for maintaining high effort. The dictator may underestimate how much effort and output depend on the incentive system; simultaneously, he may undervalue the rents he might lose to producers if monitoring were abandoned. In this case monitoring might be abandoned by mistake. However, biased expectations should then be corrected by reality, so the shift to a low state should not be irreversible: the dictator should be able to induce producers to return to a high–output equilibrium afterwards.

This plausibly interprets some of what happened under *perestroika*. Beforehand Gorbachev believed that with less monitoring producers would reduce rent–seeking and increase effort: scaling down plan assignments in his 1987 industrial reform, he expected output to rise as effort was shifted into profit–seeking, market–oriented activities. But producers equated the end of coercion with limitless possibilities for securing an income without effort. Today all know they

were mistaken, so the observed pressure from conservative political forces to reimpose controls on the economy is only to be expected; a recent survey showed 48 per cent of Russians in favour of a return to state planning and distribution; 58 per cent believe it would have been better if the country had stayed as it was before 1985.⁶¹

In fact, why hasn't it already been restored? An important reason may be that the dictator lost credibility. To restore coercion today would require the government to act tougher — to monitor more efficiently, and increase the size of punishments — compared with the old days, just to compensate for the loss of credibility.

Political shocks and loss of credibility. Suppose adverse trends in monitoring costs have eliminated the feasible space for a command economy in the long run: in figure 2, $M > m_2$. However, the dictator's commitment to high coercion retains credibility with producers who believe monitoring will never be relaxed. Monitoring remains profitable at a level between m_2 and m_3 because, despite rising costs, the dictator can pay not the minimum reward required by producers in a repeated game (condition 9) but the lower one-shot minimum (condition 6).

Think of the level of coercion as varying partly with the dictator's self-interested choices, partly with random political shocks. In the past random variation was confined within a 'high' region, but this region has shrunk. The dictator is continually tempted to capitalise on his reputation and monitor less; in the past he recognised the temptation and resisted it, but costs are rising. Then the idea of monitoring less becomes newly fashionable; a political shock reinforces temptation causing the level of coercion to stray momentarily below the high region. The unpunished strike wave of 1989 presents such an occasion. Suddenly, producers wake up: the game they thought had been played out once and for all in 1929, the year of the Great Breakthrough, is being repeated! It doesn't have to be like this forever! Monitoring can fail! Insiders may steal rents! Instantly the reward required to keep producers at high effort rises to the long-run minimum, but this, being above the maximum for profitable monitoring, is enough to make the dictator throw in the towel without putting producers through a punishment phase.

Histories of *perestroika* suggest strongly that accidents and unforced errors contributed to the loss of reputation of basic Soviet institutions.⁶² Also, the dictator may have underestimated the

fragility of his own reputation. Thus, with or without biased expectations, time–inconsistent choices may have helped precipitate the downfall of the command economy.

The redistribution of income

What light may a choice–theoretic framework cast on a process that impoverished a hundred million people? I interpret the Soviet collapse as the outcome of a struggle between the dictator and producers over the distribution of income. High effort became unsustainable because high output was no longer sufficient to cover the combined coercion costs of the dictator and effort costs of producers. Effort and output fell. In the process, both income and welfare were redistributed.

Consider welfare first. Immediately before the collapse the dictator and producers are approximately indifferent between the alternative expected payoff streams that will accrue when they remain within the high–effort equilibrium or abandon it, so for that moment we can rewrite condition 10 as an equality:

$$s - m + \mathbf{d}_D \cdot f = r = e - p + \mathbf{d}_P \cdot (s + p) \quad (11)$$

However, after the collapse, the producers will become better off and the dictator will become worse off in current welfare. The dictator will become worse off if his static losses, $s + f$, exceed the gains, $r + m$, that is:

$$s + f - m > r \quad (12)$$

which is clearly required by equation 11. Likewise producers will become better off if their static gains exceed their losses, i.e.

$$r < e + i \quad (13)$$

which again is required by equation 11.

Income is redistributed differently. The reason is that when the economy collapses output falls by more than welfare. To maintain high output, costs were incurred. These costs fell on both players: producers incurred the cost of effort, and the dictator incurred the costs of monitoring. Thus incomes exceeded payoffs when output was high, since part of the incomes of both players merely compensated them for the costs of high output. When monitoring is abandoned and effort

falls, the dictator's income will change by $r - s - f$ which, equation 11 tells us, will always be negative; producers' income changes by $s - r$ which will be negative if producers' discount factor and the effort cost of maintaining high output were sufficiently large, to be exact, if $e > (1 - d_p) \cdot (s + p)$. If income from stolen rents is unreported, as seems likely, then a fall in the reported income of producers becomes certain, while the reported income of the dictator will fall as long as $s < m + (1 - d_D) \cdot f$.

In short, we can readily explain how a command economy of the Soviet type may look stable, in fact be stable, then be induced to collapse as a result of adverse trends and a dictator's decisions that leave everyone with a lower income than before. To analyse the redistribution of producer incomes among *nomeklatura* capitalists, new Russians, and the impoverished rest would require an extension of the model to differentiate producers by their abilities to grab existing assets and create new ones. To explain a general collapse of welfare would require an extension to producers' competitive asset-grabbing, or a rise in transaction costs and loss of interindustry coordination, giving rise to further unintended consequences.⁶³

6. Conclusions

A simple model of the costs and benefits of maintaining a command system has provided parsimonious explanations for a wide range of stylised facts about the Soviet economy.

The Soviet economy was stable until it collapsed. The Soviet economy grew along a stable trend for most of the twentieth century. Production and allocation were inefficient, and welfare outcomes fell short of what was feasible under different arrangements. Long-run adverse trends in monitoring costs may have ensured that one day the Soviet economy would become unstable. However, these trends were exogenous to the command system: its eventual collapse was not a result of the Soviet economy's intrinsic properties.

Command economies can secure stable high output under given historical circumstances. Stability is conditional, and the conditions for a high-output equilibrium based on coercion can be identified. Stability means that the perceived benefits to the players in the command economy

from combining high levels of effort and coercion outweighed the costs. This does not mean everyone was better off under a command system than under well-functioning market arrangements, which have not yet emerged in Russia and are not considered here.

Coercion requires artificial incentives. An artificial incentive system is required for a high-effort equilibrium. Without such incentives, producers would avoid high effort. A dictator may efficiently offer rewards as well as penalties. From producers' point of view penalties are just negative rewards. From the dictator's point of view, efficient rewards cost more than efficient penalties. However, if feasibility or credibility limits the maximum penalty, the minimum efficient reward may be greater than zero.

Coercion can be legitimate socially but not legally. Command economies may acquire social legitimacy from efficient and credible punishments and rewards. However, authority that rests on coercion is not law-governed and cannot make commitments that are legally binding. As a result, there is a *time-consistency problem for central planners*. When punishments are costly it is optimal for a dictator to declare that coercion will never be relaxed. Producers who believe this are less likely to relax effort. When effort is less likely to be relaxed, it may become optimal for the dictator to cut coercion costs by relaxing monitoring. However, the credibility of this commitment rests on his reputation, which is fragile and may be lost if coercion is relaxed once.

In exercising coercion the dictator is rationally secretive. Both the dictator and producers may exploit information asymmetries to shift payoffs in their favour. Specifically, the dictator will conceal monitoring and punishment costs to prevent producers identifying the point of reduced effort at which he might gain from relaxing coercion; this adds to his reputation. Similarly, *under coercion producers will complain and underperform*. Producers will exploit the difficulty of observing effort to overstate its subjective costs and will engage in effort reduction to secure improved rewards.

Command economies may be undermined by adverse trends in monitoring costs. Changes in production can raise the costs of monitoring producers. When monitoring becomes unprofitable, the dictator will abandon high coercion. *Command economies may also be undermined by bad policy.* Too much and too little coercion are both destructive. Too much means overreliance on

penalties. Too little means tolerating rent-seeking and erosion of the dictator's reputation. Both can undermine the profitability of monitoring. In particular, *command economies may be undermined by socialist economic reforms*. These were designed to cut monitoring costs and enlarge the feasible space for a high-output equilibrium but probably failed to do so. Moreover, the cycle of reforms and counterreforms may have harmed the dictator's reputation.

The dictator's surrender, not workers' resistance, triggered Soviet collapse. The story told here is not one of revolution, which did not take place in Russia although it may have done elsewhere; workers' resistance informed the dictator rather than forced his hand. When adverse trends have sufficiently eroded the feasible space for a command economy, the high-output equilibrium may collapse in different ways depending on who moves first. In the Soviet case the economy collapsed because when monitoring ceased to be profitable the dictator gave up; when the dictator gave up, producers gave up too.

Endnotes

¹ For examples see McNeill (1998), Rutland (1998), Brzeski (1999), Malia (1999), Pipes (1999); the term was first applied by Dallin (1992).

² Alternative lists are offered by Pryce-Jones (1995), 19–21, and Rutland (1998), 35–7. Few were economists; Pryce-Jones nominates one (the late Peter Wiles) and Rutland none. As an economist Birman (1980) has a good case for inclusion, as does Ticktin (for copious self-references see Ticktin (1998). Khanin (1991) argued that the roots of the Soviet economy's ongoing self-destruction lay buried in past decades, and warned also that a sudden liberalisation would end in catastrophe; he correctly predicted (1992) a collapse more severe and more permanent than the Great Depression of 1929–32, but blamed it on atrocious policy mistakes rather than systemic faults. Schroeder (1995), 223–4, lists a larger number of western economists who insisted over the years on the necessity of political and ownership change for significant improvement in Soviet economic performance but these were prescriptions rather than predictions. At the time most western economists would have agreed with Millar (1987), 183: 'It would be wise to discount predictions of imminent Soviet economic collapse'.

³ Churchward (1975) and Lane (1976).

⁴ Silver (1987).

⁵ Bahry and Silver (1990).

⁶ Gibson (1993).

⁷ White (1990), Finifter and Mickiewicz (1992), Bahry (1993), Fleron (1996), Blanchflower and Freeman (1997).

⁸ On Soviet growth see Ofer (1987), Bergson (1989), Easterly and Fischer (1995), and Harrison (1998a). On measurement see Harrison (2001).

⁹ Maddison (1995), Millar and Clayton (1987), and Blanchflower and Freeman (1997).

¹⁰ Easterly and Fischer (1995), 348, 361.

-
- ¹¹ Olson (1995), Wintrobe (1998), 228–232, Grossman (1998), and Solnick (1998).
- ¹² Murphy, Shleifer, and Vyshny (1993).
- ¹³ *The Gorbachev factor* is the title of the well-known book by Brown (1997)
- ¹⁴ Kontorovich (1993), 44. In addition to Brown (1997) see also Dallin (1992), Ellman and Kontorovich (1992 and 1998), Ellman (in Treml and Ellman, 1993), Khanin (1991 and 1992), Becker (1994), and Schroeder (1995).
- ¹⁵ Dallin (1992), 296.
- ¹⁶ Brown (1997), 317.
- ¹⁷ Winiecki (1993) and Kornai (1994).
- ¹⁸ Blanchard (1997), 26–35.
- ¹⁹ On the withdrawal of subsidies see Kim (1999).
- ²⁰ On repression and loyalty see Wintrobe (1990).
- ²¹ Hunter (1961).
- ²² Wintrobe (1990).
- ²³ Mills and Rockoff (1987).
- ²⁴ Machiavelli (1961).
- ²⁵ On the ‘stationary bandit’ see Olson (1993).
- ²⁶ The ‘fractional’ Politburo revealed in 1956 by Khrushchev (1976), 76–77, originated in 1926–27 and was formalised in 1937 (Khlevniuk, 1996, 48, 237–8). On ‘Team Stalin’ see Belova and Gregory (2001).
- ²⁷ Belova and Gregory (2001).
- ²⁸ Harrison (1998b) and Markevich (2000).
- ²⁹ Belova and Gregory (2001).
- ³⁰ For illustration see the analysis of ‘hostile takeovers’ in the unofficial secondary market for motor vehicles by Gregory and Lazarev (2000)
- ³¹ Ellman (2001), 142.
- ³² Harrison (1998b).
- ³³ Dearden, Ickes, and Samuelson (1990).
- ³⁴ Weitzman (1980).
- ³⁵ Kuromiya (1988) and Davies (1989 and 1996).
- ³⁶ On the origins of forcing labour to cover enforcement costs see Khlevniuk (1993)
- ³⁷ Karklins (1989).
- ³⁸ Manning (1993).
- ³⁹ Filtzer (2000).
- ⁴⁰ Wheatcroft and Davies (1999) and Ellman (2000)
- ⁴¹ Kim (1997 and 1999).
- ⁴² Khrushchev (1976), 80–1.
- ⁴³ Fragile reputation may contribute to ‘coercion fatigue’: in the end, every dictatorship fails (suggested by Kalin Nikolov, personal communication, 23 March 1999).
- ⁴⁴ Berliner (1976), 375–80, and Harrison (1998b).
- ⁴⁵ Harrison and Simonov (2000), 237–8.

-
- ⁴⁶ V.M. Glushkov, cited by Ellman (1969), 288.
- ⁴⁷ Goskomstat (1987), 410.
- ⁴⁸ Party membership from various sources; working population from Goskomstat (1987), 411.
- ⁴⁹ Wintrobe (1998), 219–20.
- ⁵⁰ Wintrobe (1998), 225.
- ⁵¹ This account rests on Schroeder (1972, 1979, and 1982), Hanson (1983), Bornstein (1985), Brus (1986), Kornai (1986), and Kontorovich (1988).
- ⁵² Millar (1985).
- ⁵³ Olson (1995), 32.
- ⁵⁴ Gregory (1987).
- ⁵⁵ Kontorovich (1986b)
- ⁵⁶ On the social history of the strike movement see Siegelbaum and Walkowitz (1995).
- ⁵⁷ On declining plan targets see Schroeder (1985); on declining penalisation of low effort see again Kontorovich (1986b).
- ⁵⁸ Ellman and Kontorovich (1998).
- ⁵⁹ *The Observer* (London), 18 October, 1998.
- ⁶⁰ On power conversion see Mawdsley and White (2000).
- ⁶¹ *The Economist* (London), 18 December, 1999.
- ⁶² Ellman and Kontorovich (1998) and Solnick (1998).
- ⁶³ E.g. Blanchard (1997) and Solnick (1998).

References

- Bahry, Donna (1993), 'Society transformed? Rethinking the social roots of perestroika', *Slavic Review*, 52(3), 512–54.
- Bahry, Donna, and Silver, Brian D. (1990), 'Soviet citizen participation on the eve of democratization', *American Political Science Review*, 84(3), 821–47
- Becker, Abraham C. (1994), 'Intelligence fiasco or reasoned accounting?: CIA estimates of Soviet GNP', *Post-Soviet Affairs*, 10(4), 291–329
- Belova, Eugenia, and Gregory, Paul R. (2001), 'Dictators, loyal, and opportunistic agents: the Soviet archives on creating the Soviet economic system', University of Houston (available from www.soviet-archives-research.co.uk)
- Bergson, Abram (1989), *Planning and performance in socialist economies: the USSR and eastern Europe*, Boston, MA
- Berliner, Joseph S. (1976), *The innovation decision in Soviet industry*, Cambridge, MA
- Birman, Igor (1980), 'The financial crisis in the USSR', *Soviet Studies*, 32(1), 84–105
- Blanchard, Olivier (1997), *The economics of post-communist transition*, Oxford
- Blanchflower, David G., and Freeman, Richard B. (1997), 'The attitudinal legacy of communist labor relations', *Industrial and Labor Relations Review*, 50(3), 438–59
- Bornstein, Morris (1985), 'Improving the Soviet economic mechanism', *Soviet Studies*, 37(1), 1–30
- Brown, Archie (1997), *The Gorbachev factor*, Oxford
- Brus, Włodzimierz (1986), '1950 to 1953: the peak of Stalinism', '1953 to 1956: the "thaw" and the "new course"', '1956 to 1965: in search of balanced development', and '1966 to 1975: normalization and conflict', in Kaser, Michael, ed., *The economic history of eastern Europe 1919–75*, vol 3, *Institutional change within a planned economy*, Oxford, 3–249
- Brzeski, Andrzej (1999), 'The end of communist economics', in Edwards, Lee, ed., *The collapse of communism*, Stanford, CA, 119–39
- Churchward, Lloyd G. (1975), *Contemporary Soviet government*, 2nd edn, London
- Dallin, Alexander (1992), 'Causes of the collapse of the USSR', *Post-Soviet Affairs*, 8(4), 279–302
- Davies, Robert W. (1989), *The industrialisation of Soviet Russia*, 3, *The Soviet economy in turmoil, 1929–1930*, London and Basingstoke
- Davies, Robert W. (1996), *The industrialisation of Soviet Russia*, 4, *Crisis and progress in the Soviet economy, 1931–1933*, London and Basingstoke
- Dearden, James, Ickes, Barry W., and Samuelson, Larry (1990), 'To innovate or not to innovate: incentives and innovation in hierarchies', *American Economic Review*, 80(5), 1105–24
- Easterly, William, and Fischer, Stanley (1995), 'The Soviet economic decline', *World Bank Economic Review*, 9, 341–71
- Ellman, Michael (1969), *Economic reform in the Soviet Union*, London: Political and Economic Planning, PLANNING Broadsheet 509, 283–371
- Ellman, Michael (2000), 'The 1947 Soviet famine and the entitlement approach to famines', *Cambridge Journal of Economics*, 24, 603–30
- Ellman, Michael (2001), 'The road from Il'ich to Il'ich: the life and times of Anastas Ivanovich Mikoian', *Slavic Review* 60(1), 140–150
- Ellman, Michael, and Kontorovich, Vladimir (1992), eds, *The disintegration of the Soviet economic system*, London
- Ellman, Michael, and Kontorovich, Vladimir (1998), *The destruction of the Soviet economic system: an insiders' history*, London
- Filtzer, Don (2000), 'Labour discipline and criminal law in Soviet industry, 1945–1953', paper to VI World Congress of the International Council for Central & East European Studies, Tampere, Finland
- Finifter, Ada W., and Mickiewicz, Ellen (1992), 'Redefining the political system of the USSR: mass support for political change', *American Political Science Review*, 86(4), 857–74.

- Fleron, Frederic J., Jr (1996), 'Post-Soviet political culture in Russia: an assessment of recent empirical investigations', *Europe-Asia Studies*, 48(2), 225–60
- Gibson, James L. (1993), 'Perceived political freedom in the Soviet Union', *Journal of Politics*, 55(4), 936–74
- Goskomstat SSSR (1987), *Narodnoe khoziaistvo SSSR za 70 let. Iubileinyi statisticheskii sbornik*, Moscow
- Gregory, Paul R. (1987), 'Productivity, slack, and time theft in the Soviet economy', in Millar, James R., ed., *Politics, work, and daily life in the USSR: a survey of former Soviet citizens*, Cambridge, 241–75
- Gregory, Paul R., and Lazarev, Valery (2000), 'Stalin's car dealership: evidence from the Soviet state and party archives', University of Houston (available from www.soviet-archives-research.co.uk)
- Grossman, Gregory (1998), 'Subverted sovereignty: historic role of the Soviet underground', in Cohen, Stephen S., Schwartz, Andrew, and Zysman, John, eds, *The tunnel at the end of the light: privatization, business networks, and economic transformation in Russia*, Berkeley, CA: University of California, International and Area Studies Research Series no. 100, 24–50
- Hanson, Philip (1983), 'Success indicators revisited: the July 1979 decree on planning and management', *Soviet Studies*, 35(1), 1–13
- Harrison, Mark (1998a), 'Trends in Soviet labour productivity, 1928–1985: war, postwar recovery, and slowdown', *European Review of Economic History*, 2(2), 171–200
- Harrison, Mark (1998b), 'Prices, planners, and producers: an agency problem in Soviet industry, 1928–1950', *Journal of Economic History*, 58 (4), 1032–62
- Harrison, Mark (2001), 'Economic growth and slowdown', in Bacon, Edwin, and Sandle, Mark, eds, *Brezhnev reconsidered* (in preparation)
- Harrison, Mark, and Simonov, Nikolai S. (2000), 'Voenpriemka: prices, costs, and quality in defence industry', in Harrison, Mark, and Barber, John, eds. *The Soviet defence industry complex from Stalin to Khrushchev*, Basingstoke and London, 223–45
- Hunter, Holland (1961), 'Optimum tautness in development planning', *Economic Development and Cultural Change*, 9(4), 561–72
- Karklins, Rasma (1989), 'The organisation of power in Soviet labour camps', *Europe-Asia Studies*, 41(2), 276–97
- Khanin, G.I. (1991), 'Novyi etap krisiza', *Kommunist*, no. 12, 71–83
- Khanin, G.I. (1992), 'Economic growth in the 1980s', in Ellman, Michael, and Kontorovich, Vladimir, eds, *The disintegration of the Soviet economic system*, London, 73–85
- Khlevniuk, O.V. (1993), 'Prinuditel'nyi trud v ekonomike SSSR. 1929–1941 gody', *Svobodnaia mysl'*, no. 13, 73–84
- Khlevniuk, O.V. (1996), *Politbiuro. Mekhanizmy politicheskoi vlasti v 1930–e gody*, Moscow
- Khrushchev, N.S. (1976), *The secret speech delivered to the closed session of the Twentieth Congress of the Communist Party of the Soviet Union* (introduced by Zhores A. and Roy A. Medvedev), Nottingham
- Kim, Byungyeon (1997), 'Soviet household saving function', *Economics of Planning*, 30, 181–203
- Kim, Byungyeon (1999), 'Income, savings, and monetary overhang of Soviet households', *Journal of Comparative Economics*, 27(4), 644–68
- Kontorovich, Vladimir (1986), 'Discipline and growth in the Soviet economy', *Problems of Communism*, 34(6), 18–31
- Kontorovich, Vladimir (1988), 'Lessons of the 1965 Soviet economic reform', *Soviet Studies*, 40(2), 308–316
- Kontorovich, Vladimir (1993), 'The economic fallacy', *The National Interest*, 31, 35–45
- Kornai, Janos (1986), 'The Hungarian reform process: visions, hopes, and reality', *Journal of Economic Literature*, 24(4), 1687–1737
- Kornai, Janos (1994), 'Transformational recession: the main causes', *Journal of Comparative Economics*, 19(1), 39–63
- Kuromiya, Hiroaki (1988), *Stalin's industrial revolution: politics and workers, 1928–1932*, Cambridge

- Lane, David (1976), *The socialist industrial state: towards a political sociology of state socialism*, London
- Machiavelli, Niccolo (1961), *The prince*, Harmondsworth
- Maddison, Angus (1995), *Monitoring the world economy, 1820–1992*, Paris
- Malia, Martin (1999), ‘The highest stage of socialism’, in Edwards, Lee, ed., *The collapse of communism*, Stanford, CA, 71–92
- Manning, Roberta T. (1993), ‘The Soviet economic crisis of 1936–1940 and the Great Purges’, in Getty, J. Arch, and Manning, Roberta T., eds, *Stalinist terror: new perspectives*, Cambridge, 116–41
- Markevich, A.M. (2000), ‘Otraslevye narkomaty i glavki v sovetskoj ekonomike 30–ykh gg. (na primere NKTP i GUMPa)’, Russian Academy of Sciences Institute of Russian History
- Mawdsley, Evan, and White, Stephen (2000), *The Soviet elite from Lenin to Gorbachev: the Central Committee and its members, 1917–1991*, Oxford
- McNeill, Terry (1998), ‘Soviet studies and the collapse of the USSR: in defence of realism’, in Cox, Michael, ed., *Rethinking the Soviet collapse: Sovietology, the death of communism and the new Russia*, London, 51–72
- Millar, James R. (1985), ‘The Little Deal: Brezhnev’s contribution to acquisitive socialism’, *Slavic Review*, 44(4), 694–706
- Millar, James R. (1987), ‘An overview’, in Cracraft, James, ed., *The Soviet Union today: an interpretive guide*, 2nd edn, Chicago, IL, 177–90
- Millar, James, R., and Clayton, Elizabeth (1987), ‘Quality of life: subjective measures of relative satisfaction’, in Millar, James R., ed., *Politics, work, and daily life in the USSR: a survey of former Soviet citizens*, Cambridge, 31–60
- Mills, Geoffrey, and Rockoff, Hugh (1987), ‘Compliance with price controls in the United States and the United Kingdom during World War II’, *Journal of Economic History*, 47(1), 191–213
- Murphy, Kevin M., Shleifer, Andrei, and Vyshny, Robert W. (1993), ‘Why is rent-seeking so costly to growth?’, *American Economic Review Papers & Proceedings*, 83(2), 409–14
- Nove, Alec (1958), ‘The problem of “success indicators” in Soviet industry’, *Economica*, 25(97), 1–13
- Ofer, Gur (1987), ‘Soviet economic growth: 1928–1985’, *Journal of Economic Literature*, 25(4), 1767–1833
- Olson, Mancur (1993), ‘Dictatorship, democracy, and development’, *American Political Science Review*, 87(3), 567–76
- Olson, Mancur (1995), ‘The devolution of power in post-communist societies’, in Skidelsky, Robert, ed., *Russia’s stormy path to reform*, London: The Social Market Foundation, Centre for Transition Economies Paper no. 1, 9–42
- Pipes, Richard (1999), ‘The fall of the Soviet Union’, in Edwards, Lee, ed., *The collapse of communism*, Stanford, CA, 35–49
- Pryce-Jones, David (1995), *The war that never was: the fall of the Soviet empire, 1985–1991*, London
- Rutland, Peter (1998), ‘Sovietology: who got it right, and who got it wrong? And why?’, in Cox, Michael, ed., *Rethinking the Soviet collapse: Sovietology, the death of communism and the new Russia*, London, 32–50
- Schroeder, Gertrude E. (1972), ‘The “reform” of the supply system in Soviet industry’, *Soviet Studies*, 24(1), 97–119
- Schroeder, Gertrude E. (1979), ‘The Soviet economy on a treadmill of “reforms”’, in U.S. Congress, Joint Economic Committee, *Soviet economy in a time of change*, vol. 1, Washington, DC, 312–40
- Schroeder, Gertrude E. (1982), ‘Soviet economic “reform” decrees: more steps on the treadmill’, in U.S. Congress, Joint Economic Committee, *Soviet economy in the 1980s: problems and prospects*, Part 1, Washington, DC, 65–88
- Schroeder, Gertrude E. (1985), ‘The slowdown in Soviet industry, 1976–1982’, *Soviet Economy*, 1(1), 42–74
- Schroeder, Gertrude E. (1995), ‘Reflections on economic Sovietology’, *Post-Soviet Affairs*, 11(3), 197–234

- Siegelbaum, Lewis H., and Walkowitz, Daniel J. (1995), *Workers of the Donbass speak: survival and identity in the new Ukraine, 1989–1992*, Albany, NY: State University of New York Press
- Silver, Brian D. (1987), 'Political beliefs of the Soviet citizen: sources of support for regime norms', in Millar, James R., ed., *Politics, work, and daily life in the USSR: a survey of former Soviet citizens*, Cambridge, 100–41
- Solnick, Steven L. (1998), *Stealing the state: control and collapse in Soviet institutions*, Cambridge, MA
- Ticktin, Hillel H. (1998), 'Soviet studies and the collapse of the USSR: in defence of Marxism', in Cox, Michael, ed., *Rethinking the Soviet collapse: Sovietology, the death of communism and the new Russia*, London, 73–94
- Treml, Vladimir G., and Ellman, Michael (1993), 'Debate: why did the Soviet economic system collapse?', *Radio Free Europe/Radio Liberty Research Report*, 2(23), 53–8
- Weitzman, Martin L. (1980), 'The "ratchet principle" and performance incentives', *Bell Journal of Economics*, 11, 302–8
- Wheatcroft, Stephen G., and Davies, Robert W. (1999), 'The Soviet famine of 1932–3 and the crisis in agriculture', British Association for Slavonic and East European Studies Annual Conference, Cambridge
- White, Stephen (1990), *Gorbachev in power*, Cambridge
- Winiecki, Jan (1993), *Post-Soviet-type economies in transition*, Aldershot
- Wintrobe, Ronald (1990), 'The tinpot and the totalitarian: an economic theory of dictatorship', *American Political Science Review*, 84(3), 849–72
- Wintrobe, Ronald (1998), *The political economy of dictatorship*, Cambridge