Coercion, Compliance, and the Collapse of the Soviet Command Economy*

Mark Harrison**

Department of Economics
University of Warwick

Abstract

Are command systems that rest on coercion inherently unstable, and did the Soviet economy collapse for this reason? Until it collapsed, the Soviet economy did not appear unstable. Why did it then collapse? A game between a dictator and a producer shows that a high level of coercion may yield a stable high–output equilibrium, that stability may rest in part on the dictator's reputation, and that a collapse may be brought about by adverse trends in the dictator's costs and a loss of reputation. The facts of the Soviet case are consistent with a collapse that was triggered by the strike movement of 1989.

Keywords: command economy, credibility, dictatorship, incentives, monitoring, reputation. Soviet Union.

JEL classification numbers: C72, D82, P21.

This draft 4 March, 2002.

Coercion, Compliance, and the Collapse of the Soviet Command Economy

The Soviet economy began to collapse in 1990.[1] As figure 1 suggests, the suddenness with which it did so can scarcely be overstated. After rapid but turbulent economic growth under Stalin the postwar years saw nearly half a century of rarely interrupted growth. Then, Soviet real incomes fell by one third in four years. After that they fell more slowly, and during most of the 1990s incomes in Russia remained two fifths below the peak of 1989. In the middle years of the decade two fifths of the population of the Russian Federation, nearly 60 million people, lived below the official poverty line compared with two percent in the late 1980s.[2]

Existing explanations of this collapse are not entirely convincing. Among them are the propositions that the Soviet economy collapsed because it was fundamentally unstable, or alternatively because it was destabilised by one or another from a range of possible contingent factors, for example a growing gap between consumer aspirations and resources, the spread of rent–seeking and corruption, a loss of fiscal control, a loss of inter–industry coordination, and the 'Gorbachev factor'.

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: stability requires normality, and normality requires consent, but the Soviet reliance on repression crowded out consent.[3] They maintain that the essence of the Soviet system made its eventual collapse inevitable and predictable; some of them claim to have predicted it.[4]

The essentialist argument appears strong. The Soviet system was repressive, and did collapse. Yet the link from repression to collapse is usually asserted without specifics, and the mechanism 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 intellectual dissent and emigration. The dissident movement's narrow social base suggests that in the 1970s most people were not interested.[5] Surveys of the Brezhnev–era emigration found that even émigrés remained loyal to a number of basic Soviet values.[6] The Gorbachev era provides more substantial evidence of prevailing attitudes and aspirations. Surveys show that most people chose the extent of their participation in state and party institutions; the more they participated, the more influential they felt over outcomes.[7] They saw themselves as having more freedoms, with less censorship and less need for self–censorship, than many Americans and most black Americans.[8] While significant majorities favoured *perestroika* and a market economy in principle, most continued to support state ownership of heavy industry and state guarantees of basic incomes and jobs; they did not want such practical outcomes of a market economy as free prices, unemployment, or rich people.[9]

The evidence of the period refutes the idea that the Soviet economy was waiting to collapse. Time series for 1928 to 1987 show that Soviet productivity was growing.[10] It rose along a trend that was stable: the economy returned to it when subjected to a disturbance, although the latter were frequent and sometimes substantial. Trend growth became significantly slower in the mid–1970s, but remained positive. 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 extra welfare gained from the growing supply of consumer goods and services was

lessened 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.[11] Returns to Soviet accumulation were diminishing, and diminished more sharply than they should have, especially after the mid–1970s, but they remained positive.[12] The Soviet economy was slowly catching up with the advanced capitalist countries until the mid–1970s, although it was still far from overtaking them. After that its underlying growth declined further 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 it was in fact unstable the Soviet economy grew without any sign of instability for many years. Indeed one needs to 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 alternative view is that the postwar Soviet system, while not unstable at first, eventually fell victim to *a growing gap between consumer aspirations and resources*. It is often suggested that pressure on living standards from the heavy Soviet defence burden worsened this imbalance, so that military–economic rivalry with the United States added to instability. But it is hard to see such factors as decisive on their own. If the existence of a gap between households' disposable incomes and consumer aspirations was a sufficient condition for destabilisation, few economies would be left standing. And the evidence that the heavy defence burden contributed to a widening gap by damaging the growth rather than the level of living standards is not impressive.[13] Indeed it is sometimes overlooked that the Soviet defence industry was itself an important source of growth.

Another popular conjecture is that the Soviet economy eventually collapsed because of the spread of *rent–seeking and corruption*. This suggests that the Soviet system failed not because of its repressive nature but because repression failed: that is, the Soviet regime could not stop special interest groups diverting effort from production to lobbying and redistribution.[14] However, economic models of rent–seeking and corruption typically explain only poor performance and slow growth, not collapse. Exceptionally, Murphy, Shleifer, and Vyshny 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.[15] If so, it is not clear why an apparently stable regime of the Soviet type should suddenly have ceased to defend state property with such catastrophic results.

An explanation that is similar in the sense that it may make a contribution but is hardly complete in itself is the evident *loss of fiscal control* after 1985.[16] In the second half of the 1980s the budget deficit widened and almost all of it was monetised, leading to an increase in the monetary overhang.[17] In consequence the real gains to worker households from the nominal wage rises of the period resulted mainly in lengthening queues. A significant problem was accumulating. But since the authorities largely created this problem for themselves, it is not clear why they did so. In any case, the implosion of the real economy did not begin until 1990.

Some explanations for transitional recession have been based on *restrictive fiscal and monetary policies*, but these do not apply to Russia. Jan Winiecki and János Kornai have attributed output decline in central and east European transitions to demand restriction.[18] 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 supply, not demand.[19] Also placing the accent on budgetary policy, Olivier Blanchard has offered a model of output decline involving a supply shock that begins with the removal of subsidies from the state sector.[20] While it is perfectly plausible that removing subsidies adversely effected output it is difficult to apply to the Russian collapse, which began in 1990 well before subsidies were first withdrawn.[21]

A separate aspect of Blanchard's analysis that has been widely accepted as a factor in Soviet economic collapse is the sudden *loss of inter–industry coordination*; Blanchard ascribes it to a 'CMEA shock' but there was a similar shock associated with the national independence struggles of the former Soviet republics that pulled apart the Soviet trade area. In fact the sectoral and regional patterns of decline provide little support for this idea. If trade disintegration were the key factor, one would expect output decline to have been greater in the smaller, more trade–dependent states than in the relatively self–sufficient Russian republic, but the evidence for it is very weak. [22] In Russia itself one would expect that output decline was greater in the industry sectors with more highly fabricated final products, but the evidence is weaker still.[23]

Beforehand many practitioners of Sovietological economics and political science, including the present writer, were sceptical of prophecies 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.[24] Vladimir Kontorovich has written: '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'.[25]

Acceptance of a certain role for particular individuals and policies, however plausible, does not reveal what would have happened without them. At one extreme Kontorovich has asserted: 'Had Andropov lived longer or had Gorbachev turned out to be less self–assured, the Soviet system might still be with us'. More modestly, Alexander Dallin 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'.[26] Archie Brown concluded 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'.[27] Additionally these views do not explain 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.

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 some general conditions that may demarcate the command economy's 'good' and 'bad' states, and some particular circumstances that may have pushed Soviet institutions from one to the other. Part I considers the nature of economic coercion in general and the Soviet command system in particular, and identifies some costs of coercion. Part II sets the command system in an historical context of twentieth–century trends in these costs. Part III defines the players in the command system as self–interested producers and a dictator and sets out a game between the players. Part IV explores the long–run properties of the game. Part V shows that trends in variables and the actions of the players may cause a command system to collapse in various ways. In part VI this model leads us to a narrative of Soviet economic decline and collapse that is both logical and consistent with known facts. Part VII considers the inevitability, reversibility, and welfare implications of the collapse and suggests some limitations of the approach followed. Part VIII concludes.

I

In this section I consider the nature of economic coercion in general and the Soviet command system in particular, and identify some costs of coercion. I take coercion as the core relationship of a command economy. The dictator uses it to prevent the population from working for anyone else; he becomes 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.[28] Starting from this, I will develop a model that allows coercion to be set at different levels, analyses the costs and benefits, and considers the conditions for a solution in which output is high.

Coercion brings a return that diminishes. Holland Hunter argued that developing economies may gain from a degree of tension in economic planning.[29] 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 loyal by intention will be repressed.[30]

Effective coercion requires a willingness to comply, and securing compliance can be costly. Mills and Rockoff studied the regulation of food supplies in wartime Britain and the USA.[31] 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'.[32]

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 economic resources, just as repression mobilises political assets. Its effectiveness relies on the resources invested in monitoring and incentives.

Soviet coercion arose in the context of a command system: production took place in state–owned enterprises regulated by a compulsory plan. By exercising coercion the dictator established a near monopoly over capital and monopsony over labour, and was able to mobilise output. He returned part of the social product to producers for their subsistence on a line written into early Soviet constitutions but first recommended by St Paul: 'he that does not work, neither shall he eat'. Being far–sighted, the dictator aimed to maximise the long–run surplus or rent that he retained after covering his costs, and this led him to invest a considerable part of his current surplus in defending and developing the territory under his control.[33] Other costs were the resources he invested 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 extra incentives conditional upon the monitoring system. For the latter producers had to supply effort, and this supply 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 observably 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, parallel

hierarchies with overlapping spheres of responsibility, and this creates significant problems of definition. In theorising I will define the dictator and the producer by their roles: producers controlled effort, while the dictator controlled coercion. The dictator was first and foremost an individual — Stalin, Khrushchev and so on. The dictator was an economic principal who stood above the law and whose word was law. Thus planning under dictatorship was not 'law–governed'; instead, 'the plan *is* the law'. Nonetheless the dictator necessarily delegated some powers to a favoured subset of the party's Politburo; for the 1930s Eugenia Belova and Paul Gregory call this group 'Team Stalin'.[34] 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. But so far as has been learnt from the archives, these acted as the dictator's loyal agents, implementing his decisions and reporting truthfully to him on the outcomes as they saw them.[35] Thus the dictator gathered around him a number of individuals bound to his regime by a shared and encompassing interest.

Some scholars have argued that the Soviet economy's principal was not the individual dictator but a collective entity, the *nomenklatura* or party–approved list of names available for appointment to privileged state posts.[36] In this view the Soviet leader was just the self–interested, sometimes badly behaved agent of the nomenklatura. Historical research on the nomenklatura is still in its infancy, but early results suggest strongly that it was far too large, too poorly defined and factionalised in real life to be thought of as a single agent.[37]

Just as the dictator's hierarchy extended downward from above, a hierarchy of producers rose from the factory workers to managers and officials of fundholding ministries; what the latter shared was a common interest not in fulfilling the dictator's orders as such, but in tilting the ratio of the incentives they gained from the dictator relative to their effort.[38] At the apex of the system a few of the most important ministers acted both to represent the interests of producers and at the same time as insiders of the dictatorship. The resultant risk of divisions in the Politburo preoccupied Soviet leaders.[39] This is because the dictator's agents had an incentive to cheat on him when the value of an asset they could steal exceeded the value of their share in the long–run rent from that asset under the dictator.[40] The dictator had to counter the incentive to cheat by forcing his subordinates 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.[41]

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. Thus it prevented producers from stealing from the dictator. Unhindered, producers would steal inputs and products and consume them directly or trade them illicitly, and this would dissipate the dictator's rent. The possibility of stealing rents arose because the dictator had to delegate some control rights over the assets of state–owned enterprises to the managers and workers who were his agents.

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.[42] 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. We know this from evidence of systematic labour–hoarding by enterprises, combined with the fact that officials responded to suspected labour–hoarding not by increasing monitoring but by revising

incentives.[43] Effort matters when intrinsic productivities vary and the ratio of output to effort is randomised. The literature on ratcheting starts from such premisses.[44] 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.

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 necessary because, without them, producers would always prefer low effort. Soviet leaders learnt this the hard way from the command economy's formative years: in 1929 and 1930 they witnessed a vicious circle of wage equalisation, rising coercion, and declining productivity.[45] 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, although an ineffective threat against ordinary workers under conditions of a general labour shortage, powerfully threatened managers and officials who stood to lose higher pay, privileges, and promotion prospects.

Both rewards and penalties were costly to the dictator. To pay a reward the dictator had to transfer part of his rent to the producer. In enforcing a penalty he incurred both direct and collateral costs. *Direct costs* were those of enforcement, which alone could be very large: for example, at its height the Gulag employed hundreds of thousands of guards to detain, transport, and supervise forced labour. To cover such costs labour camps were treated as self–financing units where prisoners and other 'special settlers' paid for their own accommodation, subsistence, and detention out of reduced consumption.[46] However, while the dictator could shift the direct costs onto the victims, he could not do this with the *collateral costs* that arose because firing and forced labour reallocated workers to jobs of lower intrinsic productivity so that output was lost. This loss fell unavoidably on the dictator.

Historically command economies have always used both rewards and punishments. At first sight it may not be clear why a dictator should offer rewards if punishments would work as well. Call an incentive that works 'efficient'. Rewards that were efficient and resulted in the behaviour that the dictator desired had to be paid, whereas penalties that were efficient did not have to be enforced. Thus efficient penalties were cheaper than efficient rewards. This suggests that an all–powerful dictator should set penalties at whatever level is required to make them efficient subject to rewards set at zero. For example we could make our students always submit essays on time, even for zero credit, by punishing late submission with death.

In practice the dictator did not have the absolute discretion that this required. Suppose there is a maximum penalty for shirking that cannot be exceeded. If the maximum penalty is less than the efficient level, then rewards for not shirking are required as well. One upper limit is set by the potential income of which producers can be deprived; for this reason positive inducements were significant in Soviet labour camps.[47] Another upper limit is set by inherited social norms: if society expects the punishment to fit the crime, unlimited penalties for shirking may not be enforceable. 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.

For incentives to be efficient, the promise of incentives must be credible. Incentives that, when implemented, are seen to impose large costs on those who offer them may lose credibility for that reason. However, the credibility of rewards and punishments was tested with different frequency. The credibility of rewards that were large enough to be efficient was tested continually. In contrast that of punishments was only tested when they failed. If punishments were mostly efficient, their credibility was tested at correspondingly infrequent intervals.

II

Long–run trends in the dictator's costs provide a motivation for the model to be developed below. Here I suggest that a variety of influences, some domestic and some global, worked to make Soviet penalties progressively more costly, to increase the rewards required by the population to invest effort in production, and to raise the costs of monitoring the level and uses of output.

First, the *collateral costs of the penal system* are likely to have risen through time. In the 1930s and 1940s Soviet leaders experimented with very harsh penalties to combat shirking.[48] Minor production failures were treated as 'wrecking' by 'enemies of the people'. Most voluntary turnover and minor timekeeping infractions by ordinary workers were eventually criminalised regardless of individual circumstances and intentions. Those punished were commonly sentenced to forced labour in establishments subject to harsh financial constraints and physical conditions. Stalin died, and the penal system was relaxed in two stages.[49] Initially Khrushchev dismantled the system of large–scale forced labour and repealed the harsh laws criminalising petty indiscipline. Then Brezhnev increased managers' job security and greatly reduced the threat of dismissal for senior officials accused of poor performance.

An economic interpretation is as follows. Soviet human capital accumulation steadily widened the gap between the average productivities of free and forced labour. Skilled labour also became more productive relative to raw untrained labour. Thus firing and forced labour wrote off human capital of increasing value and pushed the productivity of each demoted manager or imprisoned worker down a gradient of increasing slope. Over time this steadily raised the dictator's collateral cost of reliance on extreme penalties to induce effort.

Dependence on extreme penalties also carried other kinds of collateral costs with a rising trend. For example, the rules governing the treatment of industrial absentees and quitters under Stalin were so harsh that in the late 1940s the dictator's hitherto loyal agents began to refuse to enforce them.[50] This weakened the credibility of the penal system and raised the spectre of having to intensify monitoring of the monitors in order to restore it.

The growing postwar difficulties of Stalin's penal system were not just a domestic issue. There was a global context. After World War II the victors acclaimed the outcome as a victory over fascism and exploitation. In the Cold War the superpowers increasingly competed over civil and 'human' rights. As state after state signed up to conventions that guaranteed ever higher standards of treatment of citizens by government, the Soviet Union paid a rising price for its penal system in losses of international reputation and commercial opportunities abroad.

Second, the *real rewards required to motivate effort* in the Soviet economy clearly rose through the postwar period, and probably rose faster than productivity. For example, 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.[51] Several possible reasons offer themselves. One is that more educated consumers might become less easily satisfied; Joseph Schumpeter suggested long ago that the accumulation of human capital stimulates the development of elements in society that are more interested in criticising than accumulating.[52] Another is that the Soviet economy may have overinvested in human capital.[53] Diminishing returns to rising human capital might make it difficult to maintain rewards with the result that those endowed with it might become less likely to apply it to production.

Again, any trend in the efficient reward was not a purely domestic matter. From the 1970s onwards the Soviet lag in income per head behind the United States, already large, was slowly widening, and the lag behind western Europe, although

somewhat smaller, was widening more rapidly.[54] An increasingly educated Soviet workforce was becoming more accurately informed about this gap. Comparisons with the west that were increasingly invidious may have contributed to Soviet producers' psychological devaluation of the real rewards available from their own economy.

Third, it is likely that the *costs of monitoring production* were rising too. Soviet postwar product and process innovations and structural changes promoted this trend. Product innovation brought ceaseless change in the range of products and variation of product attributes. These increased planning costs, made output 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.[55] 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 essentially unmeasured.

Process innovations associated with the transition from mass to flexible production may have raised monitoring costs by fostering producers' control over effort and information. Soviet industry had 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.[56] 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 artisan 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 direct evidence of rising monitoring costs? There was no obvious upward trend in the costs officially reported in budget outlays on planning and administration. The proportion officially engaged in 'administration' remained remarkably constant over many decades at approximately two per cent of Soviet public–sector employment.[57] This was despite fears expressed by apparently qualified observers 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'.[58]

However, important monitoring costs normally go unreported because of the hidden regulatory burden on producers. As recent experience in UK higher education indicates, hidden burdens may substantially exceed those reported in the regulator's budget.[59] One hidden burden on the Soviet economy was maintenance of the communist party's organisations in every workplace and office. These played an essential role in verifying the information and decisions of managers and officials. Producers were obliged to cover the overheads and direct costs of party facilities and activists' time. Outlays on party maintenance, if we knew them, would be a revealing proxy for the trend in Soviet monitoring costs. We do not know outlays, only membership. Membership was exclusive and controlled on the basis of 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'.[60] Overall party membership rose steadily in proportion to the adult population from less than one per cent in the early 1920s to 3 per cent in the mid–1930s, 5 per cent in the 1950s, 8 per cent in the 1970s, and 9 per cent in the late 1980s.[61] Multiplied by average hours of party work and an imputed wage these figures would suggest a similarly rising share of GDP. Thus the upward trend of party membership suggests that the monitoring of production was driven by growing needs and carried rising costs.

III

In this section I define the players in the command system as self–interested producers and a dictator and I set out a game between the players. The game shows that high coercion can produce a stable state of high output: instability is not intrinsic to a command economy. It reveals other possible equilibria including one of 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.

The dictator maximises a payoff made up by the value of rents *less* his costs and losses, $Z - M - R - S$, while the producer maximises income received in wages and bonuses and appropriated through theft, *less* costs of effort and punishments, $Y + S + R - E - P$, defining each symbol as follows:

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

The dictator and producer make choices in sequence, and both players are fully informed about each other's choices at each stage. At stage I the dictator sets coercion high or low by deciding whether or not to monitor output. When coercion is high the dictator monitors output; this prevents stealing and also enables him to promise to reward high output and penalise low output. When coercion is low the dictator is indifferent to both effort and stealing and does not monitor output: coercion is low but not zero, and the dictator still extracts a rent, but part is reallocated to producers through theft. Monitoring is costly and, when implemented, is a charge on the dictator's rent:

$$\left. \begin{array}{l} M = m \\ M = 0 \end{array} \right\} \text{ when output is } \left\{ \begin{array}{l} \text{monitored} \\ \text{not monitored.} \end{array} \right. \tag{1}$$

Monitoring output efficiently eliminates stealing. Thus, although stealing is done by producers, the dictator decides whether producers can steal:

$$\left. \begin{array}{l} S = 0 \\ S = s \end{array} \right\} \text{ when } \left\{ \begin{array}{l} M = m \\ M = 0. \end{array} \right. \tag{2}$$

At stage II the producer observes the dictator's choice and responds in that light. Depending on the promised rewards and punishments she sets effort 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, output is raised by the value of effort. Thus, controlling for the allocation of producers among employments of different intrinsic productivities,

$$\left. \begin{array}{l} X = x + f \\ X = x \end{array} \right\} \text{ when } \left\{ \begin{array}{l} E = e \\ E = 0. \end{array} \right. \tag{3}$$

At stage III, provided output has been monitored, the dictator observes the results of producers' choices made at stage II, and makes a further choice: whether or not to honour his stage–I promises to reward high output and punish low output. In the case when the dictator only makes promises that he will keep,

$$\left.\begin{array}{l} P=0 \\ P=p \\ P=0 \end{array}\right\} \text{ and } \left.\begin{array}{l} R=r \\ R=0 \\ R=0 \end{array}\right\} \text{ when } \left\{\begin{array}{l} M=m \text{ and } \left\{\begin{array}{l} E=e \\ E=0 \end{array}\right. \\ M=0. \end{array}\right. \tag{4}$$

However, if the dictator defaults on incentives $P = R = 0$ in all outcomes.

When the producer is punished for shirking there is a collateral loss of output since 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. Equation 3 was written for a given allocation of producers among employments of different intrinsic productivities, which is no longer the case when shirking is punished. In the latter case, setting $X = x$ for calibration when effort is low and unpunished:

$$\left.\begin{array}{l} X=x+f \\ X=x \\ X=x-q \end{array}\right\} \text{ when } \left\{\begin{array}{l} E=e \\ E=0 \text{ and } \left\{\begin{array}{l} P=0 \\ P=p. \end{array}\right. \end{array}\right. \tag{3a}$$

Finally, the dictator's rent, the excess of output over wages, is decided as follows, setting $Z = z$ for calibration when effort is low and unpunished:

$$\left.\begin{array}{l} Z=z+f=x-y+f \\ Z=z=x-y \\ Z=z-q=x-y-q \end{array}\right\} \text{ when } \left\{\begin{array}{l} E=e \\ E=0 \text{ and } \left\{\begin{array}{l} P=0 \\ P=p. \end{array}\right. \end{array}\right. \tag{5}$$

This framework is based on two simplifications with the aim of avoiding an unduly complex parameterisation. First, the choices facing the dictator and producer are binary rather than continuously variable. Second, I assume that the dictator and producer value each unit of current income the same; however, they will discount the future differently. The simplicity of the model that results is limiting. Designed to illustrate the conditions of a sudden collapse, it cannot be expected to shed much light on the continuous small variations of effort and monitoring that are historically observed in Soviet history before the collapse.

Figure 2 sets out the players, choices, and payoffs. Call outcome A on the far left combining high coercion with high effort a 'high' state and outcome D on the far right combining low coercion with low effort a 'low' state. In between are some mixed states, outcome B which is 'high–low' (with high coercion and low effort), and C which is 'low–high' (low coercion and high effort). Finally the high–coercion states include outcomes A' and B' in which the dictator defaults on incentive promises. But in the low–coercion states the dictator promises nothing and cannot default.

When played once the game has the following features. First, effort is costly so low effort is unambiguously the producer's best response when high effort is unrewarded and shirking unpunished. This means two cases: when the dictator does not monitor output, and when the dictator monitors output but is expected to default on promises. In each case the producer loses from high effort; the payoff to her from D is superior to that from C, and B' is superior to A'.

The dictator may try induce high effort by means of high coercion. He needs enough room within the economy's parameters that both players may gain something from high output. He has to make it possible for the producer to prefer outcome A over B, offering rewards and punishments that are credible so that her payoff becomes higher with high effort. He has to make the expected reward for effort *plus* the penalty for shirking together big enough to exceed the producer's subjective cost

of effort. Having done that he, the dictator, must still prefer outcome A over D: his return to monitoring must exceed his costs of monitoring and rewards, the return being the value of the extra effort induced *plus* the output secured from theft. In short, high coercion and high effort will pay both parties when the dictator can offer a credible reward such that $f + s - m > r > e - p$.

High coercion may still be the dictator's best strategy even if he cannot make high effort pay and he will prefer the high–coercion, low–effort outcome B to D. The condition for this is $s > m + q$. This is because monitoring does at least protect his rents from theft. The dictator's downside is the associated monitoring costs and collateral costs of penalties.

A defect of the one–shot game is that it does not allow for consequences that follow the current game or the forward–looking behaviour that anticipates them. Specifically, it fails to show why the dictator should ever keep a promise to reward or punish. This is because the dictator pays for incentives. Whether producers work hard in the expectation of rewards or shirk despite the threat of punishment the dictator always raises his short–run payoff by defaulting; for example, he prefers outcome A' to A and B' to B so that A and B are ruled out. Knowing this the producer will not find the dictator's promises credible and will always choose low effort, so that A' is ruled out too.

In fact the only outcomes available are B' and D. The dictator can choose between these two. Which is better depends on the relative costs of monitoring and not monitoring. The cost of monitoring is m. The cost of not monitoring is s, the output that the producer will steal. Thus when $m < s$ the dictator should choose B' and maintain high coercion regardless of the producer's effort choice. But if $m > s$ he will abandon monitoring and choose D.

Although the one–shot game is very restricted it does identify some possible equilibria, some conditions required to sustain monitoring in general and a high–state outcome specifically, and some different ways in which the high state might break down. The dictator will always monitor the producer provided $m < s$. A high state is the equilibrium only when incentives are credible and a reward can be set such that $f + s - m > r > e - p$. Increases in the producer's effort cost or the dictator's monitoring cost, and decreases in the value of effort or penalties and the scope for theft can disrupt this condition. So can anything that undermines the credibility of the dictator's promises. Provided the dictator's residual benefit from monitoring still exceeds his higher costs and $s > m$ the game may now end in high coercion followed by low effort, a condition closer to Poland in the 1980s than Russia in any period. Beyond a point, however, the dictator may prefer to switch to a strategy of low coercion, the economy slumping from a high state to a low state.

As a former Soviet official once 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'.[62]

IV

In this section I explore the long–run properties of the game set out above. The long run matters because the dictator's gain from keeping incentive promises only becomes apparent when the game is repeated. A repeated game offers a far–sighted dictator the possibility of a stream of rewards from high effort, but he can only secure this stream by delivering a matching stream of rewards and penalties. Promises have to be credible to be efficient, and the prospect of repetition can make his promises credible. The repeated game shows the conditions under which the dictator's best choice is to sustain a high–state equilibrium; this means it also shows conditions under which the dictator may let it collapse. The repeated game additionally shows

conditions under which the producer can force the dictator's hand and precipitate a collapse by staging a slowdown or strike.

At stage I the dictator's regime choice is conditioned by the relative costs of monitoring and not monitoring. The cost of monitoring to the dictator is m. The cost of not monitoring is now at least s, the output that the producer will steal. Take the case when $m < s$; this means the dictator's best choice at stage I is to maintain high coercion regardless of the return he expects from the producer's effort choice at stage II. Then, at stage III the dictator chooses between keeping his word and cheating on incentives; under what circumstances should a forward–looking dictator pay up? When effort is high and $m < s$ paying a reward is the dictator's best long–run strategy provided he can set r such that $\delta_D \cdot f > r > e - p$ [for proof see the appendix, **proposition 1**], where δ_D denotes the dictator's discount factor and $0 < \delta < 1$. As before, he must be able to fix the producer's net return from compliance higher than her cost of effort. This gains him a discounted stream of revenue from the high effort that is induced. If the stream is large enough both players gain from the dictator's honouring his promises, and the high state of outcome A will be the unique equilibrium. This is more likely the more far–sighted the dictator and the larger his discount factor. If the condition is not met the dictator will cancel rewards, but continue to monitor output, and the high–low outcome B', will ensue.

However, when $m > s$ monitoring is unprofitable unless the dictator's payoff is bolstered by a return from the producer's high effort. His regime choice at stage I will then be conditioned by the efficiency of incentives. Faced with the producer's effort choice at stage II, the dictator must pay up or default at stage III; if he defaults, he implicitly also chooses to move to a low–coercion regime at stage I of the following game. This is because if there is no point in paying incentives and $m > s$ there is no point in paying monitoring costs either. Under what circumstances will the dictator prefer to pay up? When $m > s$ paying a reward and continuing to monitor is the dictator's best strategy provided he can set r such that $\delta_D \cdot (f + s - m) > r > e - p$ [**proposition 2**]. Again the dictator must be able to fix the producer's reward higher than the cost of effort less the penalty that the producer avoids by working hard. This gains him a discounted stream of revenue from the producers' high effort that is induced, plus the rents secured from theft, net of his monitoring costs. Again, the larger the stream and the more far–sighted the dictator, the more likely that all players will gain from the dictator's honouring his promises, making a high state the unique long–run equilibrium. If not, the dictator will cancel rewards and cease to monitor output, and the outcome will be a low state.

Figure 3 maps propositions 1 and 2 onto a space defined by rewards on the vertical axis and monitoring costs on the horizontal axis. The result is a shaded envelope M_1 that marks out the feasible space for a high–state equilibrium: for any $m < m_1$ it is possible to set r such that a high–state equilibrium is feasible.

M_1 is bounded by two frontiers. The lower surface is an *efficiency floor*: below it, incentives are not efficient. Thus a producer who receives a reward less than $e - p$ will prefer to shirk, so the efficiency floor is a horizontal line at $r = e - p$.

The upper surface of M_1 is a *credibility ceiling*: above it, incentives are not credible. The credibility ceiling is kinked at $m = s$. From proposition 1, over the range $0 < m < s$ a dictator who has to pay a reward greater than $\delta_D \cdot f$ will prefer to cheat, but without switching regime, so the ceiling is horizontal at $r = \delta_D \cdot f$. From proposition 2, however, a dictator who has to pay a reward in excess of $\delta_D \cdot (f + s - m)$ would prefer not only to cheat but also to abandon monitoring, so at $m = s$ the credibility ceiling turns down along a line with vertical intercept at $r = \delta_D \cdot (f + s)$ and slope of $-\delta_D$.

Outside the envelope the high state will collapse; it will collapse into a high–low state to the left of $m = s$, and into a low state to the right. The intersection of the credibility and efficiency frontiers defines the maximum of m_1 to which the dictator's monitoring costs may rise before he can be induced to abandon monitoring as well as cheat. I will call the case when $m > m_1$ an m_1 violation.

What happens when a producer with incomplete information decides to test the credibility of the official punishment for low effort? The rationale for asking this question is as follows. To avoid public accountability the dictator keeps his costs and revenues secret and may also disguise his valuation of them. But this creates a credibility problem. Incentive promises are credible when meeting them, however costly, is seen to be in the dictator's interest. The credibility of efficient rewards is tested continually. But the credibility of a punishment is only tested when it fails and in a well–functioning command system this may not have happened for some time. In an atmosphere of official secretiveness and dissimulation doubts may accumulate in society as to whether the authorities remain prepared to implement penalties that are rarely enforced and potentially costly. In this situation the efficiency and credibility frontiers are redefined. We should ask what level of rewards will be efficient when producers do not expect to be punished; and what level of rewards is credible when, before the dictator can pay them again, he must first bear the cost of restoring the credibility of punishments. In short, when is it that producers who do not expect to be punished will be proved wrong? Clearly the question can only be applied to producers acting in partial ignorance.

Reconsider the producer's problem. The efficiency floor in figure 3 showed the level of reward below which producers will choose to shirk even when they expect to be punished for it. But if producers would prefer to shirk when they expect to be punished, i.e. if in figure 2 they would prefer outcome B to A, they would like outcome B' (shirk with impunity) still better, and outcome D (steal as well as shirk) best of all. If the producer could choose these she would reject a reward at the efficiency floor in figure 3. The producer's problem is how to bring these outcomes within her choice set when the dictator has the first move. Under certain conditions there is a solution and it is to stage a slowdown or strike.

In figure 4 these conditions are mapped onto the same space as figure 3. When punishments have lost credibility there is a new efficiency floor that is higher than before and steps upward at $m = s$; below it, the reward offered will not be enough to put off a strike. A new credibility ceiling, horizontal at first as in figure 3, turns down at $m = s$. Above it the reward is so high that the dictator could not be expected to gain by punishing shirkers enough to restore high effort. Within the frontiers a shaded envelope M_2 marks out the feasible space for a high–state equilibrium when punishments have lost credibility: only for $m < m_2$ is it possible to set r such that a high–state equilibrium is feasible. When $m > m_2$ I will call it an m_2 violation.

The frontiers are found as follows. First, as for figure 3 the producer's options are limited by the dictator's regime choice, and the latter are again conditioned by the relative costs of monitoring and not monitoring. When $m < s$ the dictator cannot be induced to abandon monitoring, but he can be induced to condone a strike if the cost of punishing it is high enough. To maintain a high–state equilibrium and avoid a strike when the producer does not expect to be punished it is no longer enough that $r > e - p$; the efficiency floor lifts to $r = e$. As for the credibility ceiling, it moves to

$$r = f - q \cdot \left(\frac{1}{\delta_D} - 1 \right)$$; a credible reward must be low enough that, in combination with

the collateral cost of imposing penalties, it does not exhaust the dictator's gain from restoring high output. In short the producer's best strategy is to strike and the

dictator's best response is to cancel penalties but maintain monitoring, i.e. choose outcome B', if he cannot set r such that $f - q \cdot \left(\dfrac{1}{\delta_D} - 1 \right) > r > e$ [**proposition 3**].

Where will the new credibility ceiling lie in relation to the old one? The new vertical intersect may be either lower or higher. If q is small, specifically if $q < \delta_D \cdot f$, the credibility ceiling lifts and the dictator's problem is actually made easier. The fact that the stoppage temporarily relieves him of paying a costly reward relaxes his long–run budget constraint. The dictator's strategy is to punish the strike, then reinstate rewards at a level that may be higher than before; Hungary in the 1960s or Poland in the 1970s may have resembled this outcome. If q is large, however, the cost of punishing the strike causes his credibility ceiling to descend at the same time as the efficiency floor rises, compressing his options from both sides.

When in contrast $m > s$ the dictator will abandon monitoring unless his payoff is supplemented from the producer's high effort. Faced with a strike, his regime choice is conditioned by the efficiency of incentives. In this case the producer's best strategy is to strike and the dictator's best response is to cancel penalties and abandon monitoring, i.e. choose outcome D, if he cannot set r such that

$f + s - m - q \cdot \left(\dfrac{1}{\delta_D} - 1 \right) > r > e + \delta_P \cdot s$ [**proposition 4**], where δ_P denotes the

producer's discount factor. The logic of this is that when the producer chooses to strike she not only does not expect to be punished but also takes into account her potential gain from forcing the dictator into a regime change. As a result the efficiency floor rises to $r = e + \delta_P \cdot s$; it is no longer enough that $r > e$. At the same

time the credibility ceiling turns down along the line of $r = f + s - m - q \cdot \left(\dfrac{1}{\delta_D} - 1 \right)$:

for it to be worth restoring a high–state equilibrium in face of a strike the efficient reward also must be low enough that, in combination with the dictator's immediate cost of imposing penalties and flow of monitoring costs, it does not exhaust the subsequent gain from restoring high output. Thus when there is an m_2 violation a strike can force even a long–sighted dictator who is also faced with rising costs of monitoring and penalties, or a falling value of effort, from the high state to the low state.

Proposition 4 has two further notable features. One is the general effect of the rents that may be stolen when monitoring is lifted. As s increases the incentive for producers to strike rises, but the incentive of the dictator to resist rises by more. Thus the spread of corruption opportunities, controlling for other factors, does not threaten the stability of the high state but enhances it.

Another feature is the effect of the players' different discount factors. Without a capital market to bring about convergence, the Soviet dictator discounted the future by less than producers. 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.[63] 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 through the postwar period.[64] Proposition 4 suggests that under some circumstances an excess of δ_D over δ_P helped to sustain the command system: the dictator, focused on the long term, remained willing to invest in incentives while producers, struggling to survive from day to day, remained willing to comply.

V

In this section I show that trends in variables and the actions of the players may cause a command system to collapse in various ways. Figures 3 and 4 show that a collapse could be triggered by two types of violation. The possibility of an m_1 violation is shown in figure 3: beyond a certain point the dictator would choose to cheat on rewards and abandon monitoring, and m_1 shows the maximum level of monitoring costs that the dictator could sustain before taking this option. The other possibility, shown in figure 4, is an m_2 violation: the dictator's surrender could be forced when he could no longer afford the collateral costs of an efficient penalty for shirking, and m_2 shows the maximum level to which monitoring costs could rise before producers could force him to run up a white flag.

These are mapped onto the same space in figure 5. In each diagram the vertically shaded area M_1 marks out the feasible space for a high state equilibrium when both rewards and penalties are credible. The horizontally shaded area M_2 marks out the same when producers no longer expect to be punished. As monitoring costs rise, which trigger point will the command economy encounter first? The two diagrams suggest the likely relationship between m_2 and m_1. In figure 5.1 the collateral costs of penalties to the dictator are relatively small, and M_2 is so large that $m_1 < m_2$. In figure 5.2 penalty costs are so large that M_2 is completely enclosed by M_1; this sufficiently determines $m_2 < m_1$. A comparison of the two figures suggests that m_2 remained less than m_1 for a wide range of intermediate values.[65] Therefore, taking into account the likely upward trend of punishment costs, I think of figure 5.1 as a special case and other cases where $m_2 < m_1$, including figure 5.2, as more general, at least for the late Soviet period.

When $m_2 < m_1$ it did not follow that the dictator was certain to be brought down by a strike as soon as monitoring costs passed the m_2 level. Rather, one could think of the range above m_2 and below m_1 as a zone in which the dictator gained from continued operation of the command system but could not afford to defend it if challenged. In this range, therefore, the stability of the command system could be upheld by a strategem but compromised if the strategem was exposed.

The first element in this strategem was concealment of the dictator's costs. Secretiveness was a double–edged weapon; it created doubt where the the dictator could otherwise have gained credibility from revelation, but it also served to obscure potentially dangerous truths. The policy of concealing the true maintenance costs of the monitoring and penal systems that was pursued from Stalin's time onward ensured that producers could not know at what point the dictator might self–interestedly choose to condone a strike or fail to defend the regime.

The second element was the dictator's reputation for brutality and intransigence, which combined elements of truth and bluff. It was in the dictator's interest not to be perceived as a rational actor who chooses freely and self–interestedly among available options. Thus Stalin encouraged the belief that he did not count the cost of casualties or the harsh penalisation of shirkers and deserters whether in war or in economic mobilisation, and he supported this belief through his actions. In this way he acquired the reputation of a dictator who would never condone a go–slow or strike, who would always choose high coercion, punish efficiently under all circumstances, and defend his regime regardless of the producer's effort choice and the cost of changing it. He presented himself not as a free agent but as an instrument of a deterministic History that marches ever onwards and precludes past decisions from ever being reversed. He bequeathed this reputation to his successors.

Believing in this reputation and subject to secrecy, the producer should always be satisfied with any reward above the efficiency floor of M_1, and would never strike with the hope of forcing the dictator to concede monitoring. The dictator's strategem

would permit him to sustain a level of monitoring that would be instantly compromised if the bluff were called.

VI

The framework outlined above enables us to distinguish those narratives of Soviet collapse that are logical and consistent with known facts from those that are not. First, was the dictator's reputation at issue?

The record shows that every Soviet leader until Gorbachev attached grave importance to the reputation of the regime. Each took every opportunity to uphold the permanence of the monitoring mechanism. Most famously, in 1937 Stalin said that mass terror would actually intensify as Soviet society marched nearer to communism. Two decades later Khrushchev withdrew this specific commitment but reasserted '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'.[66] Hindsight suggests that such claims protested too much, because the only thing that guaranteed them was an unbending will. If monitoring was relaxed once, the claim that it could never be relaxed would no longer be believed. The dictator's reputation was fragile: a stumble could weaken it and a major slip could destroy it.[67]

In the postwar years the process of socialist 'economic reform' undermined this reputation. Declining Soviet postwar growth with rising costs of both monitoring and not monitoring set the scene for a complex cycle of systemic reforms in the Soviet Union and eastern Europe.[68] These reforms were driven over more than two decades by the need to improve allocation while reducing monitoring costs. This was expressed as a search for a socialist economic mechanism that would regulate itself, . realigning incentives so that plans and producers could coexist with greater harmony than under traditional command arrangements. If reforms succeeded, the dictator could efficiently delegate control rights to managers without continual costly monitoring. The final attempt at such a rearrangement was Gorbachev's industrial reform of 1987.

The goal of improved allocation was an important dimension 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. Controls were 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 rising tide of regime costs.

In the cycle of reforms and counterreforms, moves 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 net 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.[69] According to Mancur Olson, state–owned enterprises became 'more nearly insider lobbies or organised special interests than productive enterprises'.[70]

In short, each reform cycle weakened the dictator's reputation. This reputation could not be rebuilt by counterreforms because it relied on an unbroken history that was being destroyed. Declining reputation and rising costs associated with monitoring the changing production system together suggest a story in which the feasible space for a high–state outcome was gradually squeezed. As in figure 5.2 the dictator could

profitably sustain monitoring costs beyond the level of m_2 on the strength of his reputation alone. As his reputation shrank, the risk grew that he might suddenly find a given level of monitoring costs that was formerly sustainable, less than m_1 though above m_2, to lie outside his feasible space.

Figure 5 illustrated two triggers for command insolvency. In either case the dictator's best response was to renege on promised incentives, but in the case of an m_2 violation the dictator was forced to abandon monitoring when no longer able to afford an efficient penalty; faced with an m_1 violation the dictator made an unforced decision to abandon monitoring when no longer able to afford an efficient reward. Which of these matches better the sequence of events that led to the collapse of the Soviet command system? The historical record tends to support a narrative of forced surrender.

A few observations from the final years of the Soviet economy are indicative of growing incentive problems, but also suggest that before Gorbachev the difficulties were capable of being put right. Under Brezhnev, economic plans became less demanding while productivity growth declined.[71] Rewards were perceived to have become unsatisfactory, and this was widely believed to contribute to productivity problems.[72] Failures of work discipline became increasingly commonplace and were unpunished. But a policy shift in 1983 under Andropov, continued by Chernenko, heightened monitoring and the penalisation of labour violations. There was also a brief reduction in repressed inflation. This led to improved effort for a few years.[73]

In short, incentives that had gradually become less efficient were temporarily made more efficient again. Not wishing to increase rewards substantially, the dictator was not afraid to raise penalties and incentive credibility was maintained. Still there is a lot that we do not know about this episode, including the scale of additional monitoring costs and collateral productivity losses. Besides, the variation of both effort and incentives through these years, although noticeable in time series, were relatively small compared with what was to follow.

Under Gorbachev producers lost their fear of penalisation and unprecedentedly went on strike for higher pay. Equally without precedent, their withdrawal of effort was not only not punished but was rewarded with concessions: the leader reneged on the repressive response to worker unrest required by Soviet tradition.[74]

Oddly enough the situation on the eve of this decisive event, shown in table 1, although disturbing, still looked far from catastrophic. Wage inflation, well controlled in 1985 and 1986, rose modestly to 8 per cent in 1988. With price controls still in place, and little productivity growth, the result was rising budget subsidies and an increase in retail shortages and queues. Accordingly the monetary overhang also increased: if prices needed to rise by 19 per cent in 1985 to clear the retail market, by 1988 the required price increase had risen to 27 per cent. Real output, however, continued to grow slowly and had not yet attained its Soviet–era peak.

In principle this pattern could be interpreted in two ways. One implication could be that the Soviet economy was experiencing an m_1 violation in which the rewards offered to producers ceased to be credible or efficient. But the fact that output continued to rise through 1989 indicates that producers did not treat it as such. It was not until the credibility of punishments was tested in the strike wave in the summer of 1989 that the collapse began.

What happened in 1989 is better described as an m_2 violation caused by a collapse of the dictator's reputation that led both producer and dictator to recalculate their options. Consider figure 5.2: suppose that monitoring costs, less than m_1 but higher than m_2 and rising, were squeezing the economy into the diminishing space between the efficiency floor and the credibility ceiling of M_1. The combination of rewards and monitoring costs already lay outside M_2 but at first producers didn't strike because of the dictator's reputation for regime intransigence and brutal penalisation of shirking. Thus far the monitoring system remained profitable for the dictator and the costs of monitoring remained affordable. Then a new leader resolved

Table 1 near here

to experiment with something new and unprecedented involving more rules and citizens' rights, wider choices, less monitoring, and less brutal punishment than before. At a certain point producers woke up: the move Stalin had played as if irreversibly in 1929, the year of the Great Breakthrough, was being replayed! The regime could change and would respond to pressure! Producers recalculated their options and set about testing the credibility of the penal system.

The unpunished strike wave of 1989 provided the confirmation they sought. The minimum reward required to keep producers at high effort rose instantly from the efficiency floor of M_1 to that of M_2. In the relevant range of monitoring costs the minimum efficient reward was above the M_2 credibility ceiling: an m_2 violation. The decisive sequence was, therefore, a producer strike followed a failure to punish, an inflationary implosion of real output, and the dismantling of the monitoring regime.

VII

In this section I consider the inevitability, reversibility, and welfare implications of the collapse and I suggest some limitations of the approach followed. In the narrative of Soviet collapse that has emerged, both deterministic trends and willful acts play intrinsic roles. The trends, all adverse, were the rising costs to the dictator of punishments, rewards, and monitoring. These alone would one day have squeezed the life out of the command system. Eventually monitoring would become a loss–making activity and would be closed down.

The adverse trends were partly contingent upon the successes of the command system. I have suggested that the costs of punishments, rewards, and monitoring were bound to rise with human capital accumulation, the limits to further exploitation of mass production, and the rise of services, together with the Soviet engagement in superpower rivalry. To this extent a deterministic process was at work that would end one day in the command system's collapse. However, it does not mean that command systems are unstable under all circumstances or may collapse at any time.

The collapse was conditioned not only by deterministic trends but also by willful acts that had unintended consequences. Players made mistakes that reflected biased expectations. These arose because only one outcome of the game was normally experienced. An inexperienced dictator inheriting a high–coercion, high–effort equilibrium might underestimate how much its maintenance relied on his reputation. He might also undervalue the rents that would be lost to shirking and theft without monitoring. These might make the dictator cancel penalties or stop monitoring too easily. 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 the unintended consequence was that effort and output fell. Another unintended consequence of his policies was the widespread public discrediting of basic Soviet institutions.[75]

Producers also probably suffered from biased expectations. Influenced by the command system's reputation for permanence, they might have colluded in a high–coercion, high–effort outcome for years beyond the point where they could have mounted a strike that would have forced the monitoring regime into insolvency. But producers might also go on strike too easily if they wrongly equated a relaxation of monitoring with limitless possibilities for securing income without effort.

If monitoring was abandoned because of mistakes arising from misperceptions, then the misperceptions should have been corrected by experience. Thus a recent survey showed 48 per cent of Russians in favour of a return to state planning and distribution; 58 per cent believed it would have been better if the country had stayed as it was before 1985.[76] Of course this too could reflect misperceptions: the evidence is that most people, including most Russians, are very poor judges of the trend in their own incomes and often tend to skew their perceptions downward.[77]

When misperceptions were corrected, why were mistaken policy choices not reversed? If the Soviet command system was made to collapse in error, why has it not been restored? One reason is that the dictator lost reputation. To restore the command system today would require the Russian government to act much tougher — to monitor more thoroughly and increase the size of punishments — compared with the old days, just to restore this reputation. A dictator might be willing to pay for this investment if the command system could still generate enough rents. But perhaps in the last years the regime was kept running on reputation alone; if reputation were lost there was not enough revenue to restore it. Thus, though the system may have been destroyed in error, deterministic trends have prevented the error from being reversed.

The framework developed in this article throws only limited light, however, on what happened after collapse and on the accompanying redistribution of income and welfare.

The coercion–effort model interprets the Soviet collapse as the outcome of a struggle between a dictator and a producer over the distribution of income. To maintain high output, costs were incurred. These costs fell on both players: the producer bore the cost of effort, and the dictator bore the costs of coercion. Incomes exceeded payoffs when output was high, since part of the incomes of both players merely compensated them for the costs of the high state. High effort became unsustainable when high output was no longer enough to cover the combined coercion costs of the dictator and effort costs of producers. When monitoring was abandoned effort and incomes fell. In the process, incomes fell by more than welfare. The dictator's income was certain to fall. Producers' true incomes were likely to fall, and their reported incomes would certainly fall if, as seems likely, their new revenues from theft were unreported. In short, it is readily shown 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 actual redistribution of producer incomes among nomenklatura capitalists, new Russians, and the impoverished rest would require extensions of the model to understand the process of power conversion.[78] Instead of a dictator and a producer, we should differentiate producers by their abilities to grab existing assets and create new ones, distinguish those elements of the dictatorship that were capable of capitalising their political associations as the endgame approached, and understand the strategy of those other elements that sought to retain political power by other means.[79] To explain a general collapse of welfare would require further extensions to analyse producers' competitive asset–grabbing as a negative–sum game, or a general increase in transaction costs and loss of inter–industry coordination, giving rise to further unintended consequences.[80]

VIII

A simple model of the costs and benefits to a dictator and a producer of maintaining a command system has supported a consistent interpretation of the trends and events leading up to Soviet economic collapse.

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 may have been feasible under different arrangements. Its economic and international successes may have conditioned those adverse trends in regime costs that ensured that one day the command system would become unstable. This does not mean that the command system was unstable by its very nature or could have collapsed at any time.

A simplified model of the command system suggests that *command economies could secure stable high output under specific historical circumstances.* Stability meant that the perceived benefits to the players in the command economy from

combining high levels of effort and coercion outweighed the costs. Stability was conditional: the dictator had to be able offer a reward for effort that conformed to limits imposed by considerations of efficiency and credibility.
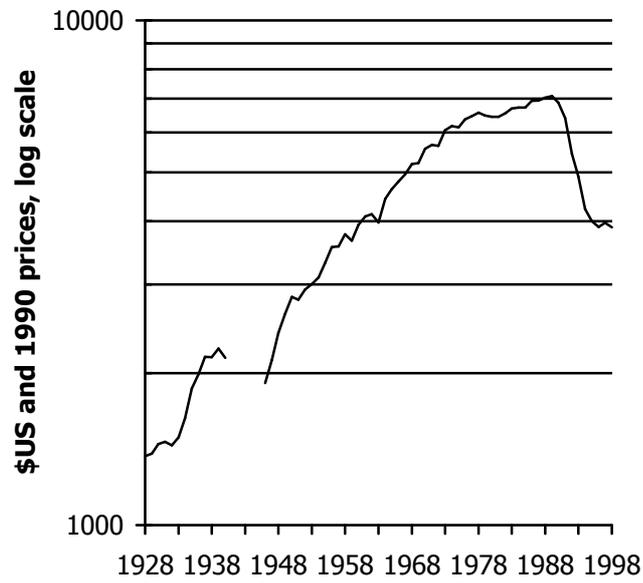
The stability of high output in a command system rested on the dictator's credibility and reputation. Under all circumstances high output relied on the credibility of the rewards and punishments promised by the dictator for high and low effort. Under specific circumstances, that is, when the collateral cost to the dictator of implementing punishments became particularly large, high output could also become reliant on the dictator's reputation as a intransigent and brutal leader.

The stability of the command system could be damaged by mistake. To the extent that a high–output equilibrium relied on the dictator's reputation, policy errors that undermined that reputation undermined the equilibrium. In particular, command economies may have been undermined by socialist economic reforms. These were designed to reduce monitoring costs but probably failed to do so. The cycle of reforms and counterreforms harmed regime credibility in the process.

The command system could follow two routes to insolvency. When adverse trends had sufficiently eroded the feasible space for a command economy, the high–output equilibrium could collapse in different ways involving the dictator's unforced or forced surrender. *The Soviet case took the form of a forced surrender*: in 1989, faced with a strike movement and no longer able to afford an efficient penalty, the dictator was compelled to abandon the monitoring regime.

Figures

Figure 1. Soviet and Former Soviet Real GDP per Head, 1928 to 1998

Source: Maddison, *Monitoring*, p. 200, and *World economy*, p. 278.
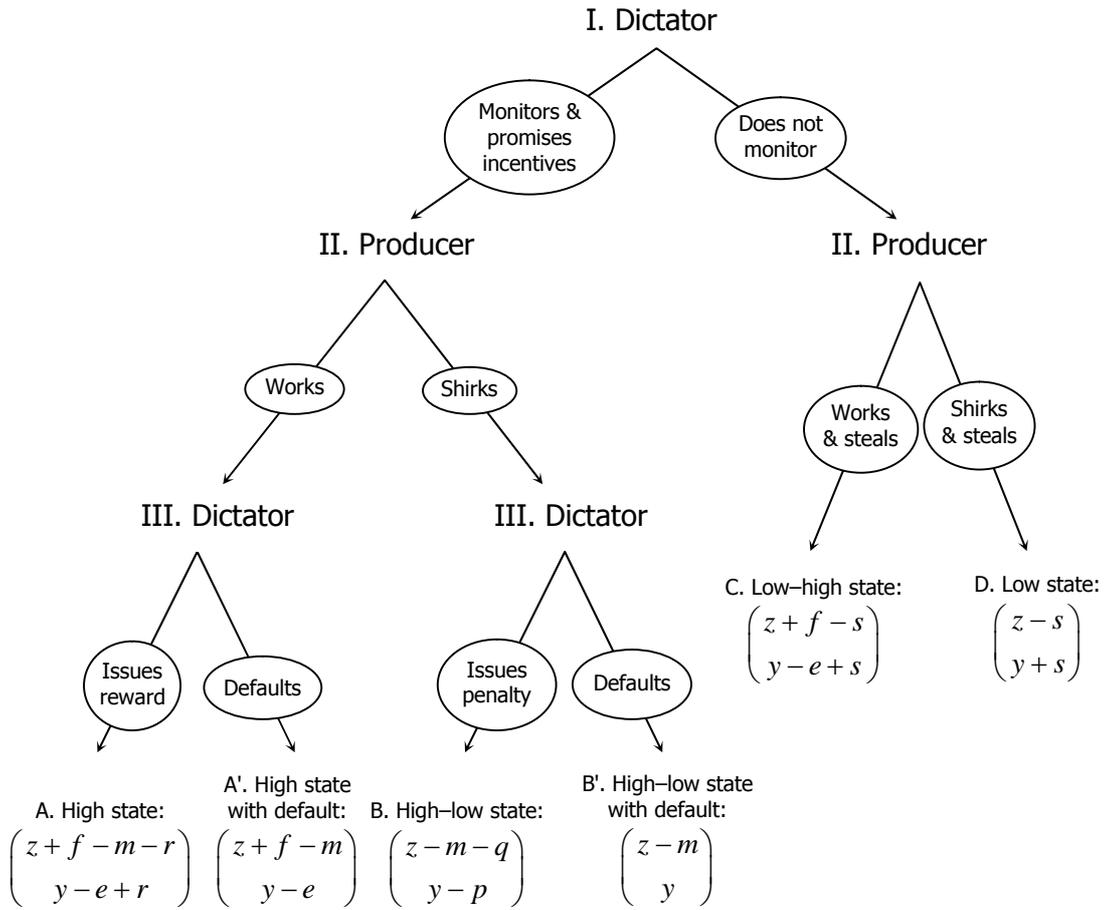
Figure 2. Players and Payoffs

I. Dictator

Monitors & promises incentives   Does not monitor

II. Producer   II. Producer

Works   Shirks   Works & steals   Shirks & steals

III. Dictator   III. Dictator

Issues reward   Defaults   Issues penalty   Defaults

C. Low–high state:
$$\begin{pmatrix} z + f - s \\ y - e + s \end{pmatrix}$$

D. Low state:
$$\begin{pmatrix} z - s \\ y + s \end{pmatrix}$$

A. High state:
$$\begin{pmatrix} z + f - m - r \\ y - e + r \end{pmatrix}$$

A'. High state with default:
$$\begin{pmatrix} z + f - m \\ y - e \end{pmatrix}$$

B. High–low state:
$$\begin{pmatrix} z - m - q \\ y - p \end{pmatrix}$$

B'. High–low state with default:
$$\begin{pmatrix} z - m \\ y \end{pmatrix}$$

Figure 3. The Efficiency and Credibility Frontiers
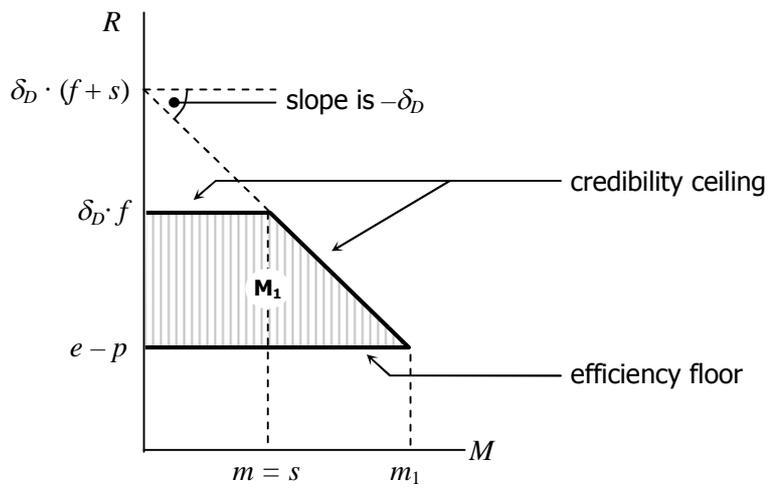
Figure 4. When Producers Expect Not To Be Punished
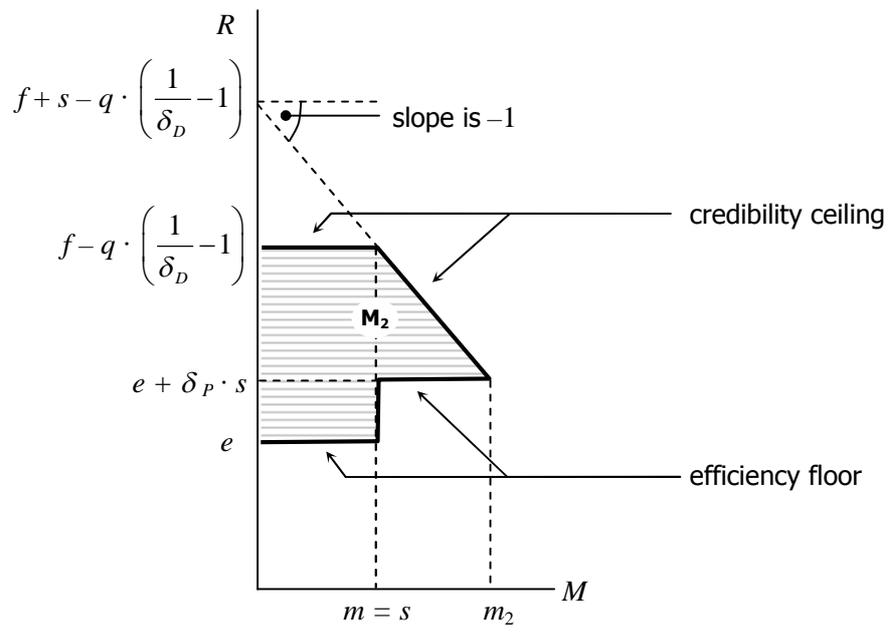
Figure 5. Triggers For Regime Insolvency

5.1. When q is very small

5.2. When q is large
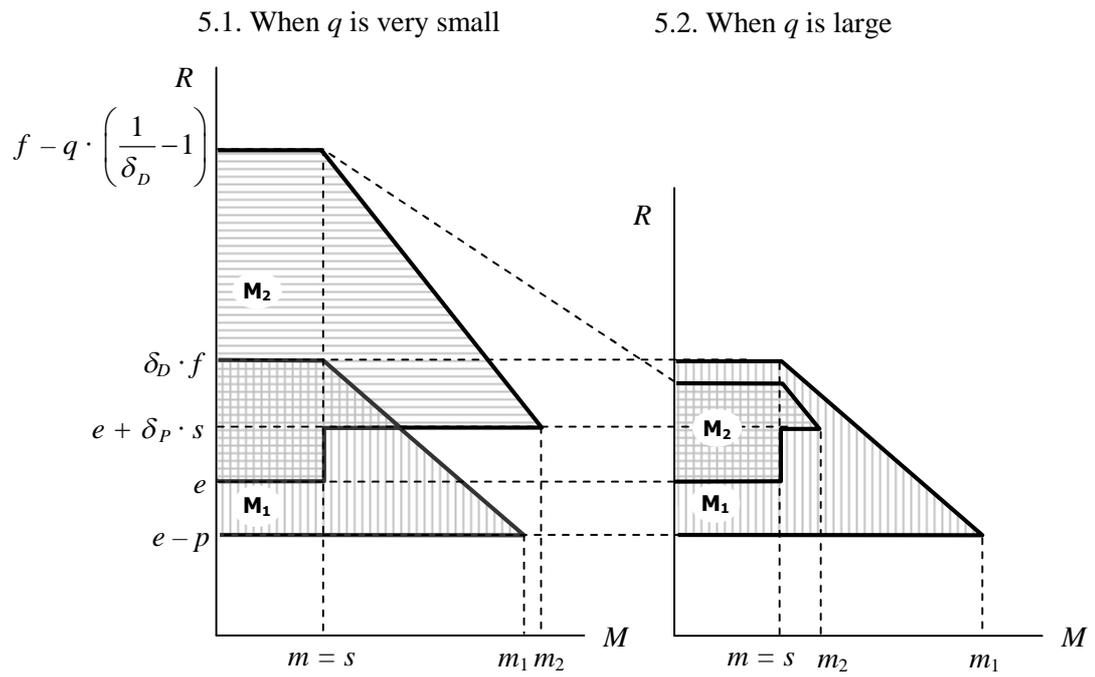
Tables

Table 1. Growing Soviet Financial Disequilibrium, 1985 to 1989

	State budget deficit, % of revenues plus deficit[a]	Average nominal public sector earnings, % change over previous year[b]	Monetary overhang, % of money stock[c]	Market–clearing retail prices, % of prevailing prices[d]	Real GDP per head, % change over previous year[e]
1985	4.6	2.9	16	119	0.0
1986	11.4	2.9	17	120	3.1
1987	13.1	3.7	19	123	0.3
1988	19.2	8.3	21	127	1.3
1989	18.6	9.4	23	130	0.7

Sources:

[a] Kim, 'Causes', table 3.

[b] Computed from TsSU SSSR, *Narkhoz 1986,* p. 397, and Goskomstat SSSR, *Narkhoz 1989*, p. 76.

[c] Kim, 'Income', p. 662.

[d] Computed as $\dfrac{1}{1-MO}$ where *MO* is the percentage monetary overhang.

[e] Computed from Maddison, *World economy*, p. 278.

Appendix

The purpose of this appendix is to prove the conditions under which the high state may be maintained in the long run assuming complete information (propositions 1 and 2) and those under which a producer who does not expect to be punished may precipitate a forced surrender by going on strike. For lower–case symbols and definitions see section 2.1 and equations 1 to 5. The dictator's and producer's discount factors are denoted by δ_D and δ_P respectively. Bold upper–case roman symbols (**A**, **B**, **C**) denote long–run payoffs as defined below.

Proposition 1. When effort is high and $m < s$ paying a reward is the dictator's best strategy provided he can set r such that $\delta_D \cdot f > r > e - p$.

Proof. Under high coercion with incentives that are credible the producer's long–run payoffs to high and low effort are $\frac{1}{1-\delta_P} \cdot (y - e + r)$ and $\frac{1}{1-\delta_P} \cdot (y - p)$; call these **A** and **B** respectively. Producers' best response is high effort provided **A** > **B**, that is $r > e - p$.

Faced with high effort the dictator must choose between **C** that denotes $\frac{1}{1-\delta_D} \cdot (z + f - m - r)$ if he pays the reward promised and two alternative payoffs:

D $= z + f - m + \frac{\delta_D}{1-\delta_D} \cdot (z - m)$ if he cheats on rewards now but continues to monitor in the future, the difference arising because if he cheats then in future games the producer will shirk; or **G** $= z + f - m + \frac{\delta_D}{1-\delta_D} \cdot (z - s)$ if he cheats now, then chooses not to monitor in the future, the difference arising because in the absence of monitoring the producer will steal as well as shirk.

When $m < s$ a high state will be the outcome provided both $r > e - p$ and **C** > **D**, that is $\delta_D \cdot f > r$.

Proposition 2. When $m > s$ paying a reward and continuing to monitor is the dictator's best strategy provided he can set r such that $\delta_D \cdot (f + s - m) > r > e - p$.

Proof. From the proof of proposition 1, when $m > s$, **D** < **G**, so a high state will be the outcome only if both $r > e - p$ and **C** > **G**, that is $\delta_D \cdot (f + s - m) > r$.

Proposition 3. Even if $r > e - p$ and rewards are credible, when $m < s$ the producer's best strategy is to strike and the dictator's best response is to cancel penalties, but maintain monitoring, if he cannot set r such that

$$f - q \cdot \left(\frac{1}{\delta_D} - 1 \right) > r > e .$$

Proof. Faced with a strike under high coercion, the dictator must choose between **H** $= z - m - q + \frac{\delta_D}{1-\delta_D} \cdot (z + f - m - r)$ from imposing penalties that, while costly to him, efficiently restore high effort in future games, and two alternative payoffs:

J $= \frac{1}{1-\delta_D} \cdot (z - m)$ if he cheats now on penalties and accepts low effort but maintains monitoring in the future, the difference arising because producers will then continue

to shirk; and $\mathbf{K} = z - m + \dfrac{\delta_D}{1-\delta_D} \cdot (z-s)$ if he not only cheats now on penalties but also chooses low coercion in future games accepting that producers will then steal as well as shirk. Thus when $m < s$ the dictator will maintain the high state provided $\mathbf{H} > \mathbf{J}$, that is $f - q \cdot \left(\dfrac{1}{\delta_D} - 1\right) > r$. However, when $m > s$, $\mathbf{J} < \mathbf{K}$ and the dictator will maintain the high state only if $\mathbf{H} > \mathbf{K}$, that is $f + s - m - q \cdot \left(\dfrac{1}{\delta_D} - 1\right) > r$.

Under high coercion when $r > e - p$ and rewards are credible the producer must choose between high effort that yields $\mathbf{A} = \dfrac{1}{1-\delta_P} \cdot (y - e + r)$ and two alternative payoffs from low effort: a strike that is unpunished yields $\mathbf{L} = \dfrac{1}{1-\delta_P} \cdot y$, and an unpunished strike followed by a shift to low coercion yields $\mathbf{N} = y + \dfrac{\delta_P}{1-\delta_P} \cdot (y + s)$.

Thus under high coercion the producer's best strategy is to strike if either of $\mathbf{L}$ and $\mathbf{N}$ is superior to $\mathbf{A}$, so long as the dictator will also accept low effort without penalty.

When $m < s$ the dictator will prefer to accept low effort without penalty if he cannot set r such that $f - q \cdot \left(\dfrac{1}{\delta_D} - 1\right) > r$ and the producer will get $\mathbf{L}$. If she is to be induced to reject the latter and prefer $\mathbf{A}$ the dictator must be able to set $r > e$; it is not enough that $r > e - p$.

Proposition 4. When $m > s$ the producer's best strategy is to strike and the dictator's best response is to cancel penalties and abandon monitoring if he cannot set r such that $f + s - m - q \cdot \left(\dfrac{1}{\delta_D} - 1\right) > r > e + \delta_P \cdot s$.

Proof. From the proof of proposition 3, when $m > s$ the dictator will prefer to accept low effort without penalty if he cannot set r such that $f + s - m - q \cdot \left(\dfrac{1}{\delta_D} - 1\right) > r$ and the producer will get $\mathbf{N}$. If she is to reject the latter and prefer $\mathbf{A}$ the dictator must be able to set $r > e + \delta_P \cdot s$; it is not enough that $r > e$.

Footnote references

Allen, R.C., 'The rise and decline of the Soviet economy', *Canadian Journal of Economics / Revue canadien d'Ēconomique*, 34(4) (2001), pp. 859–81.

Anderson, G.M., and Boettke, P.J., 'Soviet venality: a rent–seeking model of the communist state', *Public Choice*, 93(1–2) (1993), pp. 37–53.

Bahry, D., 'Society transformed? Rethinking the social roots of perestroika', *Slavic Review*, 52(3) (1993), pp. 512–54.

Bahry, D., and Silver, B.D., 'Soviet citizen participation on the eve of democratization', *American Political Science Review*, 84(3) (1990), pp. 821–47.

Baron, Samuel H., *Bloody Saturday in the Soviet Union: Novocherkassk, 1962* (Stanford, CA, 2001).

Becker, A.C., 'Intelligence fiasco or reasoned accounting?: CIA estimates of Soviet GNP', *Post–Soviet Affairs*, 10(4) (1994), pp. 291–329.

Belova, E., and Gregory, P.R., 'Dictators, loyal, and opportunistic agents: the Soviet archives on creating the Soviet economic system', *Public Choice* (forthcoming)

Bergson, A., *Productivity and the social system: the USSR and the west* (Cambridge, MA, 1978).

Bergson, A., *Planning and performance in socialist economies: the USSR and eastern Europe* (Boston, MA, 1989).

Berliner, J.S., *The innovation decision in Soviet industry* (Cambridge, MA, 1976).

Birman, I., 'The financial crisis in the USSR', *Soviet Studies*, 32(1) (1980), pp. 84–105.

Blanchard, O., *The economics of post–communist transition* (Oxford, 1997).

Blanchflower, D.G., and Freeman, R.B., 'The attitudinal legacy of communist labor relations', *Industrial and Labor Relations Review*, 50(3) (1997), pp. 438–59.

Bornstein, M., 'Improving the Soviet economic mechanism', *Soviet Studies*, 37(1) (1985), pp. 1–30.

Brown, A., *The Gorbachev factor* (Oxford, 1997).

Brus, W., '1956 to 1965: in search of balanced development', in M. Kaser, ed., *The economic history of eastern Europe 1919–75*, vol 3, *Institutional change within a planned economy* (Oxford, 1986), pp. 71-138.

Brus, W., '1966 to 1975: normalization and conflict', in M. Kaser, ed., *The economic history of eastern Europe 1919–75*, vol 3, *Institutional change within a planned economy* (Oxford, 1986), pp. 139–249.

Brzeski, A., 'The end of communist economics', in L. Edwards, ed., *The collapse of communism* (Stanford, CA, 1999), pp. 119–39.

Churchward, L.G., *Contemporary Soviet government* (2nd edn, 1975).

Dallin, A., 'Causes of the collapse of the USSR', *Post–Soviet Affairs*, 8(4) (1992), pp. 279–302.

Davies, R.W., *The industrialisation of Soviet Russia*, vol. 3, *The Soviet economy in turmoil, 1929–1930* (1989).

Davies, R.W., *The industrialisation of Soviet Russia*, vol. 4, *Crisis and progress in the Soviet economy, 1931–1933* (1996).

Dearden, J., Ickes, B.W., and Samuelson, L., 'To innovate or not to innovate: incentives and innovation in hierarchies', *American Economic Review*, 80(5) (1990), pp. 1105–24.

Deutscher, I., *The unfinished revolution: Russia, 1917–1967* (Oxford, 1967).

Easterly, W., and Fischer, S., 'The Soviet economic decline', *World Bank Economic Review*, 9 (1995), pp. 341–71.

Ellman, M., *Economic reform in the Soviet Union* (Political and Economic Planning, PLANNING Broadsheet 509, 1969), pp. 283–371.

Ellman, M., 'The social costs and consequences of the transformation process', *Economic Survey of Europe*, no. 2/3 (2000), pp. 125–40.

Ellman, M., 'The 1947 Soviet famine and the entitlement approach to famines', *Cambridge Journal of Economics*, 24 (2000), pp. 603–30.

Ellman, M. 'The road from Il'ich to Il'ich: the life and times of Anastas Ivanovich Mikoian', *Slavic Review* 60(1) (2001), pp. 140–150.

Ellman, M., and Kontorovich, V., eds, *The disintegration of the Soviet economic system* (1992).

Ellman, M., and Kontorovich, V., eds, *The destruction of the Soviet economic system: an insiders' history* (1998).

Filtzer, D., *Soviet workers and Stalinist industrialization : the formation of modern Soviet production relations, 1928–1941*, 1986

Filtzer, D., *Soviet workers and late Stalinism: labour and the restoration of the Stalinist system after World War II* (Cambridge, 2002 forthcoming).

Finifter, A. W., and Mickiewicz, E., 'Redefining the political system of the USSR: mass support for political change', *American Political Science Review*, 86(4) (1992), pp. 857–74.

Fleron, F.J., 'Post–Soviet political culture in Russia: an assessment of recent empirical investigations', *Europe–Asia Studies*, 48(2) (1996), pp. 225–60.

Gershenson, D., , and Grossman, H.I., 'Cooption and repression in the Soviet Union' (International Monetary Fund Working Paper WP/00/201, Washington, DC, 2000).

Gibson, J.L., 'Perceived political freedom in the Soviet Union', *Journal of Politics*, 55(4) (1993), pp. 936–74.

Gorlizki, Y., 'Stalin's cabinet: the Politburo and decision–making in the post–war years', *Europe–Asia Studies*, 53(2) (2001), pp. 291–312.

Graham, C., , and Pettinato, S., 'Frustrated achievers: winners, losers, and subjective well being in new market economies' (The Brookings Institution, Center on Social and Economic Dynamics, Working Paper Series no. 21 Washington, DC, 2001).

Granick, D., *Job rights in the Soviet Union: their consequences* (Cambridge, 1987).

Gregory, P.R., 'Productivity, slack, and time theft in the Soviet economy', in J.R. Millar, ed., *Politics, work, and daily life in the USSR: a survey of former Soviet citizens* (Cambridge, 1987), pp. 241–75.

Gregory, P.R., 'The dictator's orders', in P.R. Gregory, ed, *Behind the façade of Stalin's command economy* (Stanford, CA, 2001), pp. 11–33.

Gregory, P.R., and Lazarev, V., 'The wheels of a command economy: allocating Soviet vehicles', *Economic History Review* (forthcoming).

Grossman, G., 'Subverted sovereignty: historic role of the Soviet underground', in S.S. Cohen, A. Schwartz, and J. Zysman, eds, *The tunnel at the end of the light: privatization, business networks, and economic transformation in Russia* (University of California, International and Area Studies Research Series no. 100, Berkeley, CA, 1998), pp. 24–50.

Hanson, P., 'Success indicators revisited: the July 1979 decree on planning and management', *Soviet Studies*, 35(1) (1983), pp. 1–13.

Harrison, M., 'Trends in Soviet labour productivity, 1928–1985: war, postwar recovery, and slowdown', *European Review of Economic History*, 2(2) (1998), pp. 171–200.

Harrison, M., 'Prices, planners, and producers: an agency problem in Soviet industry, 1928–1950', *Journal of Economic History*, 58 (4) (1998), pp. 1032–62.

Harrison, M., 'Economic growth and slowdown', in Bacon, Edwin, and Sandle, Mark, eds, *Brezhnev reconsidered* (in preparation).

Harrison, M., , and Lockwood, B., 'What price teaching quality?', *Royal Economic Society Newsletter*, 113 (2001), pp. 3–5.

Harrison, M., and Simonov, N.S., 'Voenpriemka: prices, costs, and quality in defence industry', in M. Harrison and J.D. Barber, eds, *The Soviet defence industry complex from Stalin to Khrushchev* (2000), pp. 223–45.

Hunter, H., 'Optimum tautness in development planning', *Economic Development and Cultural Change*, 9(4) (1961), pp. 561–72.

Karklins, R., 'The organisation of power in Soviet labour camps', *Europe–Asia Studies*, 41(2) (1989), pp. 276–97.

Khanin, G.I., 'Novyi etap krisiza', *Kommunist*, no. 12 (1991), pp. 71–83.

Khanin, G.I., 'Economic growth in the 1980s', in M. Ellman and V. Kontorovich, eds, *The disintegration of the Soviet economic system* (1992), pp. 73–85.

Khlevniuk, O.V., 'Prinuditel'nyi trud v ekonomike SSSR. 1929–1941 gody', *Svobodnaia mysl'*, no. 13 (1993), pp. 73–84.

Khlevniuk, O.V., *Politbiuro. Mekhanizmy politicheskoi vlasti v 1930–e gody* (Moscow, 1996).

Khlevniuk, O.V., 'Sovetskie regional'nye rukovoditeli ot Stalina do Brezhneva: politizatsiia "nomenklatury"' (conference on 'Serving the State: Administrative Practices in the Soviet Union', École des Hautes Études en Sciences Sociales, Paris, 2001).

Khrushchev, N.S., *The secret speech delivered to the closed session of the Twentieth Congress of the Communist Party of the Soviet Union* (introduced by Z.A. and R.A. Medvedev, Nottingham, 1976).

Kim, B.–Y., 'Soviet household saving function', *Economics of Planning*, 30 (1997), pp. 181–203.

Kim, B.–Y., 'Income, savings, and monetary overhang of Soviet households', *Journal of Comparative Economics*, 27(4) (1999), pp. 644–68.

Kim, B.–Y., 'Causes of repressed inflation in the Soviet consumer market: retail price subsidies, the siphoning effect, and the budget deficit' (Bank of Finland Institute for Economies in Transition, Discussion Paper no. 9, 2000), forthcoming in the *Economic History Review*.

Kontorovich, V., 'Discipline and growth in the Soviet economy', *Problems of Communism*, 34(6) (1986), pp. 18–31.

Kontorovich, V., 'Lessons of the 1965 Soviet economic reform', *Soviet Studies*, 40(2) (1988), pp. 308–316.

Kontorovich, V., 'The economic fallacy', *The National Interest*, 31 (1993), pp. 35–45.

Kontorovich, V., 'Economists, Soviet growth slowdown and the collapse', *Europe–Asia Studies*, 53(5) (2001), pp. 675–95.

Kornai, J., 'The Hungarian reform process: visions, hopes, and reality', *Journal of Economic Literature*, 24(4) (1986), pp. 1687–1737.

Kornai, J., 'Transformational recession: the main causes', *Journal of Comparative Economics*, 19(1) (1994), pp. 39–63.

Kozlov, V.A., *Massovye besporiadki v SSSR pri Khrushcheve i Brezhneve. 1953–nachalo 1980-kh gg*. (Novosibirsk, 1999).

Kuromiya, H., *Stalin's industrial revolution: politics and workers, 1928–1932* (Cambridge, 1988).

Lane, D., *The socialist industrial state: towards a political sociology of state socialism* (1976).

Lazarev, V., 'Evolution of the Soviet elite and its post–communist transformation' (University of Warwick, Department of Economics, PERSA Working Paper No. 11, 2001).

Lewin, M., 'The Soviet nomenklatura' (conference on 'Serving the State: Administrative Practices in the Soviet Union', École des Hautes Études en Sciences Sociales, Paris, 2001).

Machiavelli, N., *The prince* (Harmondsworth, 1961).

Maddison, A., *Monitoring the world economy, 1820–1992* (Paris, 1995).

Maddison, A., *The world economy: a millenial perspective* (Paris, 2001).

Malia, M., 'The highest stage of socialism', in L. Edwards, ed., *The collapse of communism* (Stanford, CA, 1999), pp. 71–92.

Manning, R.T., 'The Soviet economic crisis of 1936–1940 and the Great Purges', in J.A. Getty and R.T. Manning, eds, *Stalinist terror: new perspectives* (Cambridge, 1993), pp. 116–41.

Markevich, A.M., 'Otraslevye narkomaty i glavki v sovetskoi ekonomike 30–ykh gg. (na primere NKTP i GUMPa)' (Russian Academy of Sciences Institute of Russian History, 2000).

Mawdsley, E., and White, S., *The Soviet elite from Lenin to Gorbachev: the Central Committee and its members, 1917–1991* (Oxford, 2000).

McNeill, T., 'Soviet studies and the collapse of the USSR: in defence of realism', in M. Cox, ed., *Rethinking the Soviet collapse: Sovietology, the death of communism and the new Russia* (1998), pp. 51–72.

Millar, J.R., 'The Little Deal: Brezhnev's contribution to acquisitive socialism', *Slavic Review*, 44(4) (1985), pp. 694–706.

Millar, J.R., 'An overview', in J. Cracraft, ed., *The Soviet Union today: an interpretive guide*, 2nd edn (Chicago, IL, 1987), pp. 177–90.

Millar, J., R., and Clayton, E., '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, 1987), pp. 31–60.

Mills, G., and Rockoff, H., 'Compliance with price controls in the United States and the United Kingdom during World War II', *Journal of Economic History*, 47(1) (1987), pp. 191–213.

Murphy, K.M., Shleifer, A., and Vyshny, R.W., 'Why is rent–seeking so costly to growth?', *American Economic Review Papers & Proceedings*, 83(2) (1993), pp. 409–14.

Nove, A., 'The problem of "success indicators" in Soviet industry', *Economica*, 25(97) (1958), pp. 1–13.

Ofer, G., 'Soviet economic growth: 1928–1985', *Journal of Economic Literature*, 25(4) (1987), pp. 1767–1833.

Ofer, G., 'Budget deficits, market disequilibrium and Soviet economic reforms', *Soviet Economy*, 5(2) (1989), 107–61.

Olson, M., 'Dictatorship, democracy, and development', *American Political Science Review*, 87(3) (1993), pp. 567–76.

Olson, M., 'The devolution of power in post–communist societies', in R. Skidelsky, ed., *Russia's stormy path to reform* (The Social Market Foundation, Centre for Transition Economies Paper no. 1, 1995), pp. 9–42.

Pipes, R., 'The fall of the Soviet Union', in L. Edwards, ed., *The collapse of communism* (Stanford, CA, 1999), pp. 35–49.

Pryce–Jones, D., *The war that never was: the fall of the Soviet empire, 1985–1991* (1995).

Rutland, P., 'Sovietology: who got it right, and who got it wrong? And why?', in M. Cox, ed., *Rethinking the Soviet collapse: Sovietology, the death of communism and the new Russia* (1998), pp. 32–50.

Schroeder, G.E., 'The "reform" of the supply system in Soviet industry', *Soviet Studies*, 24(1) (1972), pp. 97–119.

Schroeder, G.E., '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, 1979), pp. 312–40.

Schroeder, G.E., '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, 1982), pp. 65–88.

Schroeder, G.E., 'The slowdown in Soviet industry, 1976–1982', *Soviet Economy*, 1(1) (1985), pp. 42–74.

Schroeder, G.E., 'Reflections on economic Sovietology', *Post–Soviet Affairs*, 11(3) (1995), pp. 197–234.

Schumpeter, J.A., *Capitalism, socialism and democracy* (4th edn, 1954).

Siegelbaum, L.H., and Walkowitz, D.J., *Workers of the Donbass speak: survival and identity in the new Ukraine, 1989–1992* (Albany, NY, 1995).

Silver, B.D., 'Political beliefs of the Soviet citizen: sources of support for regime norms', in J.R. Millar, ed., *Politics, work, and daily life in the USSR: a survey of former Soviet citizens* (Cambridge, 1987), pp. 100–41.

Solnick, S.L., *Stealing the state: control and collapse in Soviet institutions* (Cambridge, MA, 1998).

Ticktin, H.H., 'Soviet studies and the collapse of the USSR: in defence of Marxism', in M. Cox, ed., *Rethinking the Soviet collapse: Sovietology, the death of communism and the new Russia* (1998), pp. 73–94.

Treml, V.G., and Ellman, M., 'Debate: why did the Soviet economic system collapse?', *Radio Free Europe/Radio Liberty Research Report*, 2(23) (1993), pp. 53–8.

Voslensky, M., *Nomenklatura: the Soviet ruling class* (Garden City, NY, 1984).

Weitzman, M.L., 'Soviet postwar growth and capital–labor substitutability', *American Economic Review*, 60(4) (1970), pp. 676–92.

Weitzman, M.L., 'The "ratchet principle" and performance incentives', *Bell Journal of Economics*, 11 (1980), pp. 302–8.

Wheatcroft, S.G., and Davies, R.W., 'The Soviet famine of 1932–3 and the crisis in agriculture' (British Association for Slavonic and East European Studies Annual Conference, Cambridge, 1999).

White, S., *Gorbachev in power* (Cambridge, 1990).

Winiecki, J., *Post–Soviet–type economies in transition* (Aldershot 1993).

Wintrobe, R., 'The tinpot and the totalitarian: an economic theory of dictatorship', *American Political Science Review*, 84(3) (1990), pp. 849–72.

Wintrobe, R., *The political economy of dictatorship* (Cambridge, 1998).

Official publications

CIA Directorate of Intelligence *USSR: Sharply Higher Budget Deficits Threaten Perestroyka* (SOV 88–10043U, Washington, DC, 1988).

Goskomstat Rossii, *Rossiiskii statisticheskii ezhegodnik* [abbreviated to *RSE 2000*] (Moscow 2000).

Goskomstat SSSR, *Narodnoe khoziaistvo SSSR za 70 let. Iubileinyi statisticheskii sbornik* [abbreviated to *Narkhoz za 70 let*] (Moscow 1987).

Goskomstat SSSR, *Narodnoe khoziaistvo SSSR v 1989 g. Statisticheskii ezhegodnik* [abbreviated to *Narkhoz 1989*] (Moscow 1990).

TsSU SSSR, *Narodnoe khoziaistvo SSSR v 1986 g. Statisticheskii ezhegodnik* [abbreviated to *Narkhoz 1986*] (Moscow, 1986).

World Bank, *Russian economic reform: crossing the threshold of structural change* (Washington, DC, 1992).

Endnotes

[1] Earlier drafts of this paper were presented to seminars at the University of Warwick and the London School of Economics, to the second Oxford–Houston conference on 'Initial Conditions and Russia's Transitional Economy', University of Houston (April 2001), and in nontechnical versions 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. I thank all the participants for their comments, especially Leonid Borodkin, Stephen Broadberry, Nick Crafts, Bob Davies, Nick Feltovich, Don Filtzer, Peter Law, Valery Lazarev, and Stephen White. Thanks also to Michael Ellman, Gregory Grossman, Philip Hanson, Byung–Yeon Kim, Grigorii Khanin, Vladimir Kontorovich, Kalin Nikolov, Robert Skidelsky, and the anonymous referees for advice.

[2] Ellman, 'Social costs', p. 126.

[3] For examples see McNeill, 'Soviet studies', Rutland, 'Sovietology', Brzeski, 'End of communist economics', Malia, 'Highest stage', Pipes, 'Fall of the Soviet Union'; 'essentialists' were first so called by Dallin, 'Causes'.

[4] Pryce–Jones, *War that never was*, pp. 19–21, and Rutland, 'Sovietology', have proposed alternative lists of scholars who predicted the collapse. Few of them were economists: Pryce–Jones mentions only one, the late Peter Wiles, and Rutland none. As an economist Birman has a good case for inclusion; see his 'Financial crisis' of 1980. So does Ticktin, based on his own 'Soviet studies'. In 1991 Khanin, 'Novyi etap', 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; in his 'Economic growth' of 1992 he correctly predicted 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. In 'Reflections', pp. 223–4, Schroeder lists a 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, 'Overview', p. 183: 'It would be wise to discount predictions of imminent Soviet economic collapse'.

[5] Churchward, *Contemporary Soviet government*, and Lane, *Socialist industrial state*.

[6] Silver, 'Political beliefs'.

[7] Bahry and Silver, 'Soviet citizen participation'.

[8] Gibson, 'Perceived political freedom'.

[9] White, *Gorbachev*, Finifter and Mickiewicz, 'Redefining the political system', Bahry, 'Society transformed?', Fleron, 'Post–Soviet political culture', Blanchflower and Freeman, 'Attitudinal legacy'.

[10] On Soviet growth see Ofer, 'Soviet economic growth', Bergson, *Planning*', Easterly and Fischer, 'Soviet economic decline', Harrison, 'Trends', and Maddison, *World economy*. On measurement see Harrison, 'Economic growth'.

[11] Millar and Clayton, 'Quality of life'; Blanchflower and Freeman, 'Attitudinal legacy'.

[12] The causes of postwar slowdown in Soviet growth are still debated. Explanations, not all mutually exclusive, include diminishing technical progress (Bergson, *Productivity*), lack of scope for capital–labour substitution (Weitzman,

'Soviet postwar growth' and Easterly and Fischer, 'Soviet economic decline'), the end of postwar recovery (Harrison, 'Trends'), falling plan tension (Schroeder, 'Slowdown'), and rising misallocation (Allen, 'Rise and decline').

[13] Easterly and Fischer, 'Soviet economic decline', pp. 348, 361.

[14] Anderson and Boettke, 'Soviet venality', Olson, 'Dictatorship', Wintrobe, *Political economy*, pp. 228–232, Grossman, 'Subverted sovereignty', and Solnick, *Stealing*.

[15] Murphy, Shleifer, and Vyshny, 'Why is rent–seeking'. Solnick, *Stealing*, studies Soviet state collapse as a slide into a 'bad' equilibrium.

[16] CIA Directorate of Intelligence, *USSR*; Ofer, 'Budget deficits'.

[17] Kim, 'Income'.

[18] Winiecki, *Post–Soviet–type economie*; Kornai, 'Transformational recession'.

[19] World Bank, *Russian economic reform*, p. 13.

[20] Blanchard, *Economics*, pp. 26–35.

[21] On the withdrawal of subsidies see Kim, 'Causes'.

[22] The change in real GDP per head in various years compared with 1990 can be regressed on the log of 1990 population across 15 former Soviet republics; data are from Maddison, *World economy*, pp. 340–1. The estimated slope coefficient for each year after 1990, while always positive as suggested by the Blanchard hypothesis, fails a 10 per cent significance test over every interval except that which extends to 1992.

[23] The change in real output compared with 1990 across 17 industry branches in the Russian Federation (data from Goskomstat Rossii, *RSE 2000*, p. 302), can be regressed on a dummy variable of 'fabricatedness' that I set to 0 for extractive sectors (oil, gas, and coal mining, ferrous and nonferrous metallurgy, the timber and paper industry, and the industry for construction materials), 1 for processing sectors (electricity generation, oil refining, chemicals and petrochemicals, and six branches of the light and food industries), and 2 for engineering and metalworking. Again the estimated slope coefficient for each year after 1990, while consistently negative as expected, fails a 10 per cent significance test over every interval for which figures are available.

[24] *The Gorbachev factor* is the title of the well–known book by Brown.

[25] Kontorovich, 'The economic fallacy', p. 44. In addition to Brown, *Gorbachev factor*, see also Dallin, 'Causes', Ellman and Kontorovich, *Disintegration*, and idem, *Destruction*; Ellman in Treml and Ellman, 'Debate'; Khanin, 'Novyi etap', and idem, 'Economic growth'; Becker, 'Intelligence fiasco'; and Schroeder. 'Reflections'.

[26] Kontorovich, 'Economists', p. 676; Dallin, 'Causes', p. 296.

[27] Brown, *Gorbachev factor*, p. 317 (my emphasis).

[28] On repression and loyalty see Wintrobe, 'Tinpot and totalitarian'.

[29] Hunter, 'Optimum tautness'.

[30] Wintrobe, 'Tinpot and totalitarian'.

[31] Mills and Rockoff, 'Compliance'.

[32] Machiavelli, *The prince*.

[33] On the economics of proprietary dictatorship see Olson, 'Dictatorship'.

[34] The 'fractional' Politburo revealed in 1956 by Khrushchev, *Secret speech*, pp. 76–77, originated in 1926–27 and was formalised in 1937 (Khlevniuk, *Politbiuro*, pp.

48, 237–8); on its postwar forms see Gorlizki, 'Stalin's cabinet'. On 'Team Stalin' see Belova and Gregory, 'Dictators'.

[35] Gregory, 'Dictator's orders'; Belova and Gregory, 'Dictators'.

[36] Voslensky, *Nomenklatura*, Gershenson and Grossman, 'Cooption', Lazarev, 'Evolution'.

[37] Khlevniuk, 'Sovetskie regional'nye rukovoditeli'; Lewin, 'Soviet nomenklatura'. According to Lewin there were separate nomenklaturas at every level and regional division of the party and state. In 1946 that of the party central committee alone comprised more than 40 000 positions; summing all the lower level and regional nomenklaturas might generate a million posts.

[38] Harrison, 'Prices'; Markevich, 'Otraslevye narkomaty'.

[39] Belova and Gregory, 'Dictators'.

[40] For illustration see the analysis of 'hostile takeovers' in the unofficial secondary market for motor vehicles by Gregory and Lazarev, 'Wheels'.

[41] Ellman, 'Road from Il'ich to Il'ich', p. 142.

[42] Harrison, 'Prices'.

[43] Dearden, Ickes, and Samuelson, 'To innovate'.

[44] Weitzman, '"Ratchet principle"'.

[45] Kuromiya, *Stalin's industrial revolution*; Davies, *Soviet economy*; idem, *Crisis and progress*. This phase ended with Stalin's attack on wage levelling which he called 'petty–bourgeois egalitarianism', followed by the rehabilitation of payment by results and large skill and seniority premia.

[46] On the origins of forcing labour to cover enforcement costs see Khlevniuk, 'Prinuditel'nyi trud'.

[47] Karklins, 'Organisation'.

[48] Filtzer, *Soviet workers and Stalinist industrialization*; Manning, Soviet economic crisis'.

[49] Kontorovich, 'Discipline', p. 21.

[50] Filtzer, *Soviet workers and late Stalinism*, pp. 158–200.

[51] Gregory, 'Productivity'.

[52] Schumpeter, *Capitalism*, pp. 145–55. Schumpeter's argument was made in relation to the intellectual under capitalism, but seems potentially applicable to any society that promotes education. Isaac Deutscher, *Unfinished revolution*, shared this belief, predicting that Soviet educational policies would eventually undermine the basis of bureaucratic rule.

[53] The economic literature that analyses Soviet growth in terms of a CES aggregate production function suggests that the Soviet economy overinvested in capital generally (for example Easterly and Fischer, 'Soviet economic decline'). David Granick, *Job rights*, pp. 202–34, has suggested that the Soviet economy overinvested in human capital specifically.

[54] Maddison, *World economy*.

[55] Berliner, *Innovation decision*, pp. 375–80, and Harrison, 'Prices'.

[56] Harrison and Simonov, 'Voenpriemka', pp. 237–8.

[57] Goskomstat SSSR, *Narkhoz za 70 let*, p. 410.

[58] V.M. Glushkov, cited by Ellman, *Economic reform*, p. 288.

---

[59] Harrison and Lockwood, 'What price'.

[60] Wintrobe, *Political economy*, pp. 219–20.

[61] Gershenson and Grossman, 'Cooption', p. 19. Taking into account numbers of pensionable age, 9 per cent of the adult Soviet population in the late 1980s was approximately 15 percent of the working population. Gershenson and Grossman analyse rising party membership as the outcome of a shift by the nomenklatura between strategies of repression and cooption. Lazarev, 'Evolution', analyses the same as the outcome of a Ponzi–type pyramid scheme operated by the nomenklatura.

[62] *The Observer*, 18 October, 1998.

[63] Wheatcroft and Davies, 'Soviet famine', and Ellman, '1947 Soviet famine', show that famine processes were also partly attributable to unintended consequences of policy.

[64] Kim, 'Soviet household saving function'; idem, 'Income'.

[65] To be precise, $m_1 < m_2$ requires that $q < e - \dfrac{1}{1-\delta_D} \cdot \left(p - \delta_D \cdot \delta_P \cdot s \right)$; bear in mind that probably $0 < \delta_P < \delta_D < 1$ while $\dfrac{1}{1-\delta_D} > 1$ and was also potentially a large number, so the condition is relatively restrictive.

[66] Khrushchev, *Secret speech*, pp. 80–1.

[67] Fragile reputation might contribute to 'coercion fatigue': in the end, every dictatorship fails (Kalin Nikolov, personal communication, 23 March 1999). Fragile reputation also reinforces Machiavelli's dictum that the most dangerous moment for a regime is when it begins to reform itself.

[68] This account rests on Schroeder, '"Reform"'; idem, 'Soviet economy'; idem, 'Soviet economic "reform" decrees'; Hanson, 'Success indicators revisited'; Bornstein, 'Improving', Brus, '1956 to 1965', and idem, '1966 to 1975'; Kornai, 'Hungarian reform process'; and Kontorovich, 'Lessons'.

[69] Millar, 'Little Deal'.

[70] Olson, 'Devolution', p. 32.

[71] Schroeder, 'Slowdown'.

[72] Gregory, 'Productivity'.

[73] Kontorovich, 'Discipline'; Kim, 'Income'.

[74] On the social history of the 1989 strikes see Siegelbaum and Walkowitz, *Workers*. On traditional responses see Kozlov, *Massovye besporiadki*, and for the treatment meted out to strikers in Novocherkassk in 1962 Baron, *Bloody Saturday*.

[75] Ellman and Kontorovich, *Destruction*, and Solnick, *Stealing*.

[76] *The Economist*, 18 December, 1999.

[77] In a large sample of Russian households used by Graham and Pettinato, 'Frustrated achievers', nearly three quarters of those whose incomes doubled or more between 1995 and 1998 assessed their income change negatively, whereas three quarters of those who lost income made accurate assessments.

[78] 'Power conversion' is the term used by Mawdsley and White, *Soviet elite*, for the private capitalisation of party networks and enterprises that accompanied the collapse of the command system.

[79] On elite continuity in Russia in transition see Lazarev, 'Evolution'.

[80] Blanchard, *Economics*, and Solnick, *Stealing*.