Накопление и принуждение к труду в период «позднего» сталинизма

Пол Р. Грегори
Университет Houston
США

Марк Харрисон
Университет Warwick
Великобритания

Тезисы
Как показывает изучение довоенных архивных материалов, Сталин стремился максимизировать количество ресурсов, направляемых на накопление. Главным ограничением такой политики являлось требование сохранить потребление, прежде всего рабочих, на достаточном для эффективного стимулирования их производственных усилий уровне. В докладе изложена простая модель сталинской экономической политики, определявшей одновременно объемы продукции, производственных усилий, реальной зарплаты, и накопления. Значительное влияние на размер зарплаты, за которую рабочие были готовы предлагать свои усилия, оказывал уровень принуждения к труду.

С 1940 года и до смерти Сталина такое принуждение держалось на высоком уровне с помощью драконовского законодательства, согласно которому даже незначительные нарушения карались заключением в ГУЛАГ. Успешное применение принуждения предполагало уменьшение ресурсов, выделяемых на потребление, без ущерба для количества затраченных рабочими производственных усилий и выпускаемой продукции.

Однако нельзя применять сплошное принуждение без ошибок. Во-первых, на применение суровых карательных мер советские рабочие и служащие реагировали перераспределением своих усилий от производства в пользу самообороны, через укрепление горизонтальных и вертикальных неформальных отношений и взаимной поддержки. Во-вторых, атмосфера принуждения увеличивала вероятность ошибок при распределении труда и капитала. В-третьих, сталинские руководители часто толковали попытки подчиненных защититься посредством «знакомств и связей» как доказательство сознательной или несознательной измены партии и Сталину, и на этом основании сами организовывали новые политические репрессии.

В таких условиях возможно принуждение не способствовало увеличению уровня накопления. Авторы полагают, что результаты, демонстрируемые советским народным хозяйством, улучшились, после того как наследники Сталина снизили уровень принуждения к труду.

Date of draft: 17 November 2004
Accumulation and Labor Coercion Under Late Stalinism

Paul R. Gregory
University of Houston
USA

Mark Harrison
University of Warwick
Great Britain

The archives tell us that Stalin was obsessed with accumulation, which is hardly a surprise. One of the great stylized facts about the Soviet economy under Stalin was the sharp increase in its accumulation rate, accompanied by the fierce repression of private consumption. Between 1928 and 1937 Stalin shifted nearly 30 percent of Soviet GNP out of private consumption into investment and government purchases; the investment rate doubled (Bergson 1961, pp. 217, 237). More generally, investment rates in socialist countries were consistently higher than in capitalist countries of comparable size and income levels (Kuznets 1963, pp. 352-57). Olson (1993) argued that a self-interested proprietary dictator or sedentary bandit who settles on and monopolizes a territory would maximize his rent through time by taxing heavily and reinvesting the surplus.

The traditional literature said nothing about how Soviet investment and consumption rates were actually chosen for the simple reason that little was known about it. Bergson’s authoritative Economics of Soviet Planning (1964) introduces investment first in Chapter 11 and largely puts aside the issue of its total volume, focusing instead on rational choice of individual projects. Bergson recognized (1964, p. 241) that desired investment depends “on the output mix to be produced from it”; this view percolated through the various editions of Gregory and Stuart (1974, pp. 214-5). The most widely cited paper on investment (Grossman 1953) deals exclusively with the choice to be made when a given increase in desired output could be obtained from alternative projects.

There is a considerable literature on how a socialist economy ought to determine aggregate investment; it stems from Marx’s emphasis on accumulation as the source of economic growth and proceeds via the growth models of Preobrazhensky and Feldman in the early Soviet period (Erlich 1960; Spulber 1964) to the later models of Brus, Kalecki, and Dobb (Rutland 1985, pp.103-107). Explicitly or implicitly these models suggested that the optimal investment rate depended on society’s discount rate, to which Preobrazhensky added the scope for redistribution away from “former” social classes.

No Politburo member was trained in economics, however. Their advisers in Gosplan were focused, at best, on balancing supplies and demands for broad
commodity groups no more than one year ahead. When Stalin and the Politburo thought of the longer term, it was probably not in terms of intertemporal optimizing; their language consisted of “urgently,” “with maximum speed,” “by [a given] date,” and so on. They had to balance multiple objectives that included holding power and eliminating enemies against economic goals. At best, we would expect that they resorted to rules of thumb. These rules should tell us not only what they hoped to achieve but also how they experienced and perceived the objective limits on their power to achieve it.

1. The Politburo Accumulation Model

At the core of the Politburo’s strategy to “build socialism” in the first two five-year plans were massive programs for the hydroelectric dams, machinery complexes, vehicle works, blast furnaces, railways, and canals that were included on its “title lists” of approved projects. Their purpose was to embody the new society in cement and structural steel. How were these programs implemented and managed?

Although Politburo meetings for the 1930s left few formal minutes, his deputies wrote to Stalin from time to time to detail key Politburo meetings and seek approval of operational decrees (Gregory 2004, chaps 4 and 5). They reveal that the Politburo consistently set three targets in the 1930s: the nominal investment budget, grain collections, and foreign exchange. These three control variables all related to investment. The investment budget allotted funds to industrial and regional agencies for construction and machinery. Despite objections to the lack of material balances for investment goods, Ministries and regional authorities were simply given “investment rubles,” and no one appeared to know the real investment that resulted. Grain collections were designed to contribute to national savings through the excess of sale prices over purchase prices. Stalin personally directed foreign exchange to the import of capital goods rather than the luxury goods sometimes demanded by the Bolshevik elite.

If Stalin’s goal was indeed to maximize investment, the archives provide two types of evidence that are, at first glance, confusing. First, Stalin was extremely concerned about consumption, particularly where it touched upon the productivity and morale of the industrial workers. Consumer supplies were one of the most frequent items on Politburo agendas; in Stalin’s words the “provisioning of workers” was one of “the most contested issues” before the Politburo and trade was “the most complicated ministry” (quoted by Gregory 2004, pp. 93-94). Stalin interpreted declining labor productivity as a sign that workers “were not provisioned as well as last year,” and personally ordered the delivery of consumer goods to cities where labor productivity was declining (Gregory 2004, chapter 4). The Politburo set itself up as the highest trading organization, deciding retail trade plans, prices, assortment, and even the opening of new stores. Second, we have the little-known empirical fact that the Politburo, which based the success of its program on capital construction, deliberately reduced nominal investment on two occasions, in 1933 and 1937 (Davies 2001b) citing the fear of inflation and rising construction costs.
This evidence can be interpreted in two ways. One is that Stalin’s preferences were unstable; fitful humanitarian impulses led him to direct more supplies to the workers from time to time; periodic fits of financial orthodoxy led him to regret overstraining the economy. Alternatively, Stalin’s preferences were stable, and he reallocated consumption or reduced investment according to a consistent rule. Knowing of Stalin’s capacity for calculation, patience, and self-control, we reject the first explanation and investigate further the second.

Figure 1. The Politburo Model

Figure 1 illustrates the model that the Politburo appears to have used to set investment and consumption. The figure mirrors the Marxian concept of the surplus product, the gap between output and consumption, as the outcome of a distributive struggle. The model has theoretical precursors in Schrettl (1982, 1984) and Podkaminer (1989), and is set out more fully by Gregory (2004, chap. 4). It belongs to a general class of models in which a ruler’s freedom of action is circumscribed by social “tolerance limits” (Kornai 1980, pp. 211-14) or a revolution or disorder constraint (Acemoglu and Robinson 2000). By raising investment without limit the Politburo risked provoking the workers to go slow, strike, or rebel.

In the Soviet economy the demand for labor was always enough for full employment and besides all able-bodied persons were required to work by law (Granick, 1987), so the figure takes employment, $N$, as fixed exogenously; individual effort, $e$, was variable, so total effort, $E = e \cdot \bar{N}$, was variable although employment was not. Total output, $Q$, depended on total effort, $E$. Total effort varied with the real wage, $w$, as follows. The aggregate wage bill $W$, the consumer goods received by workers, is measured along the vertical axis in the same units as output, and is proportional to the real wage given that employment is fixed, i.e. $W = w \cdot \bar{N}$. There is a reservation wage, analogous to a tolerance limit or disorder constraint, below which effort is zero; there is also
a “fair” wage at which effort is maximized. 1 As the economy moves from the fair wage to the reservation wage, effort declines as workers turn up late or drunk, or go slow or absent without leave; at the limit unrest simmers and threatens to boil over into local and general strikes and rebellion. Thus the effort curve intersects the horizontal axis at the reservation wage and becomes vertical at the fair wage. Effort also depends on the level of direct coercion, $C$, to which we turn in the next section. To maximize effort the dictator would pay the fair wage and get the maximum output, but this would not maximize the surplus. To maximize the surplus, $Q - W$, he would choose the intermediate wage, effort, and output levels denoted $W^*, E^*$, and $Q^*$.

The effort curve makes the consequences of plan mistakes asymmetric: for the dictator, paying the workers too little is much worse than paying them too much. Starting from the investment optimizing position, too much consumption does at least raise effort and output, and this mitigates the fall in investment. A mistake of the same size in the opposite direction, however, not only cuts effort and output by more but also risks pushing the workers into outright confrontation with the state. An investment-maximizing dictator must tread a fine line between the pursuit of investment and the triggering of serious disorder.

As noted above, Stalin clearly understood the link between effort and consumption. Stalin managed worker morale and effort in two ways. When investment and consumption were about right in aggregate, detailed plan mistakes could still leave some workers with too much and others with too little. When this happened, it was imperative to reallocate supplies immediately to those left short, and Stalin personally ordered the reallocation of consumer goods to cities where labor productivity was declining. But there could also be aggregate plan mistakes; when too much investment threatened to disrupt the economy and provoke the workers, the Politburo preferred to sacrifice investment. This is how we interpret the unforced investment cutbacks of 1933 and 1937. Although Stalin did not express his concern for consumption directly, when he advocated “strengthening the ruble” as a justification for less investment (Gregory 2004, pp. 236-242) it was the same ruble that workers had to spend in retail markets; when he feared rising construction costs, these were fiscal costs that had to be covered through taxation or the inflation tax. In this sense Stalin’s behavior was stable and consistent, given the constraints under which he operated.

Figure 1 lays out other options and dangers. Team Stalin could seek to manipulate the effort curve by offering ideological rewards in place of material payoffs. Idealists from Russia and abroad in fact assisted the first five year plan, motivated by the idea of building socialism for future generations. The attempt to transform *homo economicus* into *homo sovieticus* led, however, to a vicious circle of wage equalization, and declining productivity (Kuromiya 1988; Davies 1989a, 1996). Subsequent mobilizations were limited to short

---

1 The fair-wage concept in Figure 1 can be related to the observation of Akerlof (1984) and Akerlof and Yellen (1990) that workers in market economies give extra effort if they are paid a wage perceived to be fair, and withdraw it if the wage is reduced below the psychologically determined fair level. We suppose that effort has a maximum, so payments above the fair level do not elicit more effort.
lived campaigns such as for World War II and to cultivate the “virgin lands” of Kazakhstan and Siberia after the war. The most publicized mobilization campaign of the 1930s, the Stakhanovite movement, was driven by progressive piece rates that permitted participating workers to drive up their incomes by overfulfilling norms (Davies and Khlevnyuk 2002); The archives suggest that Stalin abandoned it because it caused widespread upset to notions of a fair wage, and also threatened inflation (Gregory 2004, pp. 104-106).

Stalin could also elicit extra effort in the present by promising future promotion. The problem was that these promises eventually had to be made good. Lazarev (2005) has identified the creation of new posts that carried a wage premium to reward loyal supporters as an important feature of Stalin’s policy. Relentlessly purging the potentially disloyal made it cheaper for him to keep his promises to the actually loyal.

In an extension of the model, the Politburo could bear down on consumption selectively. Rather than spread the burden equally, and invite widespread worker unrest, Stalin tried to confine the negative consequences to lower-priority workers by introducing ration entitlements differentiated by category and occupation according to the principle: “He who does not work on industrialization shall not eat” (cited by Gregory 2004, p. 98: emphasis added). It is notable that Stalin himself wrote the rules and procedures of the rationing system that remained in effect until 1935.

Scholars have taken different views of rationing. Osokina’s (2001) archival study argues that Stalin positively wanted to drive down consumption and introduce a discriminatory ration. Hessler (2004) argues on the contrary that Stalin approved of non-rationed trade and a consumption oriented lifestyle for the workers but crises were so frequent that this goal was only achieved for short periods. Gregory (2004, pp.98-100) argues that Stalin saw rationing as a way to force accumulation without a loss of effort of high-priority workers, despite occasional pro-consumer statements.

After largely eliminating rationing in 1935, periodic crises -- pre-war rearmament, wartime mobilization, and postwar famine -- caused restorations of rationing. Davies and Khlevnyuk (1999) and Hessler (2004) have documented the drawbacks of widespread rationing for the regime: illegal markets spread rapidly and drew off the energy of the working population; officials charged with determining individual entitlements were paralysed by a flood of petitions and paperwork; local officials and managers charged with distributing goods in practice had to search, bargain, and trade, or even steal to trade on their own account. Productive effort and budget revenues fell away. Gregory (2004, pp. 98-100) argues that Stalin abandoned rationing because the resulting distribution did not get consumer goods to priority workers.

It is not surprising, therefore, that the Politburo was typically planning to restore “normal” retail as soon as possible. Characteristically, as legal markets reappeared, Stalin sent the secret police into them to confiscate windfall gains of market traders, whom he regarded as class enemies (Hessler 2004).

Figure 1 has a further ominous extension. The fair wage was set by a mass psychology that was unpredictable and hard to manipulate. If workers concluded from the propaganda of economic successes that they were being cheated the fair wage would rise, forcing the Politburo to cut investment back. Stalin used the vast informant network of the NKVD to monitor protests, strikes, anti-Soviet statements, and factory-wall graffiti, and eavesdrop on
private conversations to gauge mass opinion (Berelovich 2000). Stalin had obvious political motives to do this, but within our framework we conclude that wages and fairness lay at the cross-hairs of politics and economics.

2. Coercion: Three Experiments

In principle efficient penalties are cheaper than efficient rewards since “a promise is costly when it succeeds and a threat is costly when it fails” (Schelling 1960, p. 177). In Figure 1, the fear of punishment may cause workers to lower their reservation wage without reducing effort; coercion may not make effort more productive but should make people willing to supply effort for less. As long as coercion displaces the effort curve downward while leaving the production curve undisturbed, the surplus is increased. In market economies, outside options, such as alternate employment, leisure, or unemployment, cannot easily be limited by force as the ubiquity of informal sector employment and illegal migration demonstrates. In the Stalinist state, however, the idea of controlling workers’ alternatives by force proved attractive.

The archives show that Stalin believed that a wide range of problems could be solved by force. The language with which he addressed both administrative and economic problems was habitually violent. Faced with foot-dragging in the transport ministry in 1931, he wrote: “we must smash this gang … If you can manage without my help, smash the gang before it’s too late”. Considering the role of speculators in a legalized urban rural market, he wrote in 1932: “we must eradicate this scum … The OGPU [secret police] and its agencies must, without delay, start training its forces and studying the enemy” (Davies et al. 2003, pp. 95, 102, 165).

Some observers have concluded that coercion was the glue that held the Stalinist economy together. When coercion failed, or was abandoned, the economy fell apart (Kontorovich 1986; Blanchard and Kremer 1997; Harrison 2002). Others have emphasized the huge social and economic costs of coercion (Conquest 1987; Khlevnyuk 2001b; Davies and Wheatcroft 2003). The balance of the argument is yet to be struck. Whether coercion paid for itself, and where the optimum lay, are unresolved issues.

Stalin conducted three notable experiments with the use of coercion to foster accumulation: the forced collectivization of the peasantry, the criminalization of workplace indiscipline, and the widespread use of forced labor. In each case the outlines have been known for decades, but the full consequences are only now coming to light and we review them below.

Stalin’s belief in solving problems by force also contributed to the motivation behind the Great Terror of 1937/38, but in this case his purpose was not to foster accumulation. Manning (1993) speculated that Stalin deployed terror to solve widespread economic problems, but the archives have not yielded any substantial evidence to this effect (Davies 2004). Rather the aim of the terror was political: to isolate and remove a potential “fifth column” from society that, Stalin believed, could endanger his regime in a time of rising international tension (Khlevniuk 1995).
2.1. Collectivizing Peasants

Politically, collectivization aimed to impose Soviet power in the countryside and eliminate the stratum of richer peasants, the kulaks. It was triggered, however, by a grain marketing crisis that reflected the peasants’ perceived unwillingness to contribute sufficiently to investment-led industrialization. Collectivization began in earnest in December 1929, signalling Stalin’s victory in the power struggle. The relatively open discussion in the late 1920s and other published sources supported a substantial literature on the collectivization decision, the most complete account of which is Davies (1980); Heinzen (2004a) has now added a study of the agriculture ministry in the 1920s. Jasny (1949), Lewin (1968), and Strauss (1969) outlined some general results. On the basis of estimates by Soviet historians Millar (1974) and Ellman (1975) suggested that the investment surplus that Stalin hoped to gain from collective agriculture was probably not forthcoming. But the detailed consequences of collectivization and the mass deportation or detention of peasants who were excluded or resisted were concealed behind a thick veil of secrecy.

Tauger (1991), Davies, Tauger, and Wheatcroft (1995), Wheatcroft and Davies (2002), and Davies and Wheatcroft (2003) have now reviewed the archives on the immediate aftermath of collectivization, including the famine of 1932-33. The collective farms enabled Moscow to replace local decision making with its own detailed plans, instructions, and rules. Stalin was focused on what he could control: sown acreage, and the state’s share in what this acreage produced. But Stalin could not control the harvest. While yields collapsed, the share delivered to the state increased. Excessive procurements, bad weather, plan errors, and rural shortages combined to strip the countryside of grain; first the livestock were killed, then the farmers themselves starved. They were prevented from feeding themselves from their own harvests by severe punishments including death for petty theft from collective farm stocks. Davies and Wheatcroft dispel Conquest’s (1987) notion that Stalin manufactured the famine to kill class enemies; rather they show the leadership subsequently trying to ameliorate the effects of its own bungling.

Famines usually occur in poor countries with limited statistical reporting. The Soviet archives have provided unusually good documentation of the two peacetime famines of 1932/33 and 1946/47. Ellman (2000) has applied the entitlement theory of Sen (1983) to the latter famine, which particularly affected the peasants of the Ukraine and other southern regions. Sen argued from historical experience that famines are more likely to arise from entitlement or distributional failures than production failures, and that the state should generally play a substantial role in cushioning sections of the population threatened by entitlement failures. Ellman makes two points from the evidence.

First, he argues, in the 1946/47 famine, production and entitlement failures interacted. The production failure was not such as to leave insufficient food available to keep everyone alive. In this sense the famine arose from a failure of entitlement. But the famine was triggered by a harvest failure; if the harvest had not failed, everyone would have had sufficient entitlements to stay alive. Therefore, the famine cannot be attributed to entitlement failures alone. Second, Ellman notes that in this famine the role of the state was essentially negative: it selected those who died by denying them entitlements. Therefore,
it cannot be a general rule to concentrate grain stocks in the hands of the state in times of famine. The motivations of the state must also be considered. Under some circumstances, by implication, market allocation may do better than the state at keeping people alive. While Ellman does not extend his findings to the Soviet famine of 1932/33, to do so would seem reasonable to us; Davies and Wheatercroft (2003, p. 402) concur at least on the first issue, the interaction of entitlements with harvest failure.

2.2. Regimenting State Employees

As the 1940s began, Stalin redirected coercion from specific class enemies to the entire public-sector work force. A battery of intimidating laws criminalized work place violations which had previously been managed by administrative sanctions within the enterprise. The laws themselves were not secret and were described in detail by Conquest (1967, pp. 102-6) and Nove (1969, pp. 260-265). Their consequences, however, remained obscure until the archives became available. Full texts of the relevant decrees and other official documents are to be published by Danielson et al. (2005, vol. 1).

The notorious law of 26 June 1940 lengthened the normal working day and week for all state employees. At the same time absenteeism, defined as any 20 minutes’ unauthorized absence during the working day, or even idling on the job, also became a criminal offense; the penalty was set at up to six months’ corrective labor at work with a 25 percent reduction in pay. Three offenses in one month or four in two months counted as unauthorized quitting, punishable by two to four months’ imprisonment. Enterprises managers were made criminally liable for failure to report worker violations. In August 1940 the minimum sentence for petty theft at work and “hooliganism” was set at one year’s imprisonment. In December 1941 defense industry workers were conscripted at work under a quasi-military discipline that prescribed 5 to 8 years in a labor camp for unauthorized quitting. Further measures of April and May 1943 militarized transport workers and required unauthorized quitters to serve 3 to 10 years in labor camps. After the war a decree of June 1947 set the minimum sentence for any theft of state or socialized property at five to seven years imprisonment. Transport work was eventually demilitarized in March 1948 and work in the defense industry two months later; otherwise, these punitive laws remained on the books until Stalin died.

A report prepared as background for Khushchev’s secret de-Stalinization speech of February 1956 (Danielson et al. 2005, vol. 1, statistical appendix), shows that from 1940 through June 1955 the regular courts and military, transport, and labor camp tribunals sentenced 35.8 million persons in total for all offenses. Not allowing for repeat offenders, this would represent about one third of the adult population of roughly 100 million. Of the 35.8 million, 15.1 million were imprisoned and a quarter of a million were executed. The annual rate of imprisonment was one million or more in most years up to 1950, and more than half a million at the time of Stalin’s death. These totals were dominated by workplace offenses. Between 1940 and 1952 there were 15 million convictions for absenteeism and unauthorized quitting, out of a total public sector workforce in 1940 of 31 million. There were 11 million sentences to corrective labor at work with loss of pay, more than 2.7 million prison terms of a few months, and nearly one million terms in labor camps of between 3 and 10 years (Filtzer 2002, pp. 159-67; Sokolov 2003, 2005).
were also 4.7 million convictions for theft of socialized property (Danielson et al. 2005, vol. 1).

The archives show that Team Stalin, not an independent judiciary, determined the number of convictions. When the laws were enacted, Stalin had to intervene personally to force a reluctant judiciary to prosecute tardy or lazy workers (Sokolov 2003).

Given the many factors that differentiated the Soviet judicial and penal system from others, it is hard to find appropriate standards for detailed comparison of workplace offending. It is simpler to go to the aggregate of all offenses. From 1940 to June 1955 the USSR imprisoned 15.1 million people or, on an annual basis, roughly one percent of its adult population each year. In 1958, for comparison, the United States imprisoned not more than one fifth of one percent of its adult population. Thus Stalin imprisoned his subjects at least five times more frequently than the United States.

2.3. The Labor Camps
Collectivization in the early 1930s and the repression of state employees in the 1940s created huge flows into labor camps. In the 1920s there was just one Soviet forced labor complex in the Arctic where mainly political detainees were held. The first major expansion came with collectivization which threw hundreds of thousands of peasants into camps or special settlements; the Gulag, the interior ministry’s chief administration of labor camps, was created in 1930 to handle the sudden inflow and labor camps spread across the remote interior and the far north and east to house, at their peak, more than 2.5 million inmates.

Although the Gulag was shrouded in intense secrecy, its human side leaked out through the recollections of former inmates. Estimates of its scale and scope, however, could be based on little more than guesswork until the first official figures were released in 1989. While contemporary estimates of the number of detainees ranged up to 20 million, we now know that there were approximately 2.3 million penal laborers at the outbreak of World War II and about 2.5 million on Stalin’s death; similar numbers of deportees were also confined to labor settlements in the remote interior. The forced laborers were mostly engaged in forestry, mining, and construction, where they made up substantial shares of employment, but never more than about 3 percent of the total workforce including farm workers, and less than this in terms of the value of national output. Bacon (1994) compares the figures reported in the archives with the range of estimates that went before. But these new facts must be set beside others including the very high rates of conviction and sentencing that we now know about: although the Gulag population was smaller than observers had sometimes guessed, it also had much higher turnover with very large numbers entering the system for relatively short periods before returning to society. Possibly, the higher estimates for numbers of forced laborers

---

2 See U.S. Department of Commerce (1960, pp. 139, 141). U.S. statistics of the time did not give aggregate sentencing figures. Instead, we multiplied the number of known crimes committed (p. 139), by the proportion of offenses cleared by arrest and the proportion of arrests that resulted in convictions (p. 142). This calculation does not allow for multiple offending.
reflected the 20-million plus cumulative flow rather than the much smaller stock at a point in time (Danielson et al. 2005, vol. 1, general introduction).

Internal Gulag documents confirm that political strategy, not economics, dictated the Gulag’s development. The labor camps were populated as a consequence of collectivization, terror, draconic labor laws, and also the punishment of ethnic groups, prisoners of war, and others suspected of potential or actual wartime disloyalty. Before the archives there was speculation that in the course of the 1930s economics eventually took over from politics as a motivation for recruiting forced laborers, or alternatively that the NKVD became a lobby for forced labor (Swianiewicz 1965, pp. 161-2). Neither of these has turned out to be the case, although our judgement has to be carefully shaded.

The Gulag had a consistent economic raison d’être: to explore and colonize regions that were resource-rich but inhospitable, since forced labor could be ordered around the country at will. Penal labor was to cover its own costs, which were to be held down to a low subsistence minimum (Khlevnyuk 2003a). Repression of the forced laborers’ wages combined with the enforcement of effort through close supervision and monitoring were supposed to promote low-cost accumulation. At the same time, no one was planning to lock everybody up for economic reasons.

At the level of the state, the continued primacy of politics over economics is shown by the fact that the NKVD-MVD did not lobby for expansion. The NKVD projected a shrinking number of Gulag inmates for the third five-year plan (1938-42), just as the first victims of Stalin’s Great Terror began to flood in (Gregory 2003a, p. 4). After the war the NKVD again planned for contraction and politics again overturned this expectation (Khlevnyuk 2003a, p. 52). In the late 1940s Gulag officials proposed to release all but the most dangerous prisoners from camps (Tikhonov 2003), but this was unacceptable to Stalin. In 1953, within three months of Stalin’s death in March, MVD chief Lavrenty Beriia had released one and a half million prisoners, 60 percent of the Gulag’s inmates. Tikhonov (2003) shows that this move had been prepared five years earlier. In fact, elements within the MVD were increasingly alarmed by its economic and social costs. The economic costs were reflected in its growing financial deficits; the social costs were measured by high rates of recidivism. By Stalin’s death, at least one quarter of Soviet adults had been through the Gulag for one reason or another (Danielson et al. 2005, vol. 1, general introduction). Although the camps were supposed to segregate hardened criminals from petty youth offenders, the camp population was a mixing bowl and recidivism soared. The high turnover spread the culture and mores of camp life throughout society.

Why did the Gulag fail? Research on the microeconomics of the Gulag is in its infancy but early case studies show a complex learning process. The Far Eastern camps (Nordlander 2003) highlighted the perennial tension between the possibilities for exploiting the forced laborers and the requirements of maintaining them. The Karelian camps (Joyce 2003) show the experimental process by which the authorities learnt the scope and limits of the exploitation of forced labor. The White Sea-Baltic Canal project (Morukov 2003) was initially seen as a success and this stimulated illusory expectations for the future. The Gulag leaders willingly undertook the building of Noril’sk (Borodkin and Ertz 2003) because they underestimated the risks and
difficulties that would arise. The operation of Noril’sk (Ertz 2003) helped to
expose illusions about the ease with which the inmates could be manipulated
and coerced into supplying effort.

By the postwar years Gulag officials believed they knew why the camps
failed to generate a surplus. Labor productivity there was extremely low
relative to that of free workers, while guarding detainees was very expensive;
in 1950 there was one guard to ten inmates, leading to the widespread practice
of “unguarded” prison contingents. Within the Gulag, prisoners formed
protective networks among themselves and with the guards to cover for each
other (Heinzen 2004b). Camp managers learned that an arsenal of
punishments was not sufficient to motivate prisoners and that there were
complicated tradeoffs: prisoners placed on reduced rations for failing to meet
work quotas were no longer able to work effectively. One of the most effective
incentive systems, early release for exemplary work, deprived the Gulag of its
best workers. Material incentives played an ever larger role in motivating
penal labor (Borodkin and Ertz 2003, 2005; Ertz 2005). In the last years of the
Gulag there was a process of “conversion of slaves to serfs” (Khlevnyuk
2003a, p. 57); the camps increasingly paid prisoners civilian wages and the
distinctions between penal and free labor became blurred.

3. Coercion Failure?
The collective farms ceased to be an instrument of coercion in the mid-1960s
when the farmers were placed on fixed wages like any other employee. The
Gulag was emptied of mass prisoners between 1953 and 1957. The draconian
labor laws of the 1940s were also rescinded in the mid-1950s. These
relaxations have two possible interpretations: Either the post-Stalin leadership
did not have the stomach for continued repression and abandoned a system
that was working; or coercion was abandoned because it did not work. The
new evidence supports the second interpretation, particularly the fact that the
most coercive instrument, the Gulag, was dissolved by its own leaders, who
included men such as Beria renowned for their amorality.

In short, the initial belief that results could be obtained more cheaply by
raising the penalty for failure than by increasing the reward for success
appears to have been mistaken. Effective coercion requires that penalties be
accurately assessed and targeted, and that the agents of repression are well
informed about offenders and the costs of their crimes. It proved more difficult
than they expected to target penalties accurately; as a result the relationship
between true effort and punishment was “noisy.” The devising of an efficient
coercion system is then further complicated if workers and managers respond
strategically to increase the noise.

The archives show that officials had little idea whether workers were
exerting full effort or not; the law could do little more than ensure that they
were physically at work and did not steal too much. Agricultural controllers
could order the collective farms to sow more land but could not assess whether
the land was being farmed efficiently (Davies and Wheatcroft 2003). In
industry, attempts to pin “normal” effort down to objective technological
criteria proved fruitless; attempts to maintain effort norms when productivity
should have been rising often gave rise to damaging social conflicts (Davies
Because the state relied on indirect indicators of effort, mistakes crept into the effort/punishment relationship. The investigation of low effort could yield an error of Type I that punished the innocent, and a Type II error that acquitted the guilty. Errors of both types appear to have been prevalent. Numerous Type I errors are reflected in the very high rates of penalization that we have reported; these condemned hard workers along with neer-do-wells, drunks, and thieves. The law criminalized such a wide range of behaviors that their scope made virtually every worker liable to prosecution for something, including one-time and accidental violations: at this time, a broken-down commuter bus could make criminals of scores of hapless workers at a stroke. Rational managers might wish to select the truly guilty for prosecution, the problem workers and repeat offenders, but the laws subjected even petty offenses to harsh penalties and managers who did not seek them in every case were threatened with the same. As a result, the innocent were bundled through the courts and camps along with the guilty in extraordinarily large numbers.

Memoirs suggest that Team Stalin probably knew this but did not care. In March 1937 Stalin’s chief instrument of the Great Terror Nikolai Ezhov told his officials to expect “some innocent victims … Better that ten innocent people should suffer than one spy get away. When you chop wood, chips fly” (Montefiore 2003, p. 194). Stalin’s prime minister, Molotov (interviewed by Chuev 1991, p. 416) also cared more about condemning the guilty than acquitting the innocent, “never mind if extra heads fall.”

Type II errors were also clearly numerous. The evidence for this is the surprising fact that, although penalization rates were very high, offending rates were even higher. Filtzer (2002, pp. 167-8) used the records of eight production branch ministries in 1947 to show that almost one third of a million instances of “labor desertion” gave rise to fewer than 55,000 convictions for a conviction rate of 16 percent. In other words, a judicial system that made the chips fly somehow failed to chop the wood. The combination of severe penalization with low conviction probability for the guilty is consistent with high-cost policing and justice administration (Becker 1968, p. 184); the high rate of conviction of innocents, however, is more properly seen as a cost of dictatorship (Djankov et al. 2003), in the sense that the dictator’s efforts to achieve a lower rate of offending than society was willing to tolerate had highly suboptimal results.

The effectiveness of the Politburo accumulation model rested on the dictator’s ability to create a gap between the civilian wage as a “fair” return for effort, and low subsistence in the Gulag as the return to shirking, so that the difference between them was the intended punishment for shirking. The widespread mistakes that we have described subverted this strategy: Type I errors reduced the worker’s expected return to effort while Type II errors raised the expected return to shirking.

Error rates were not exogenous. They were fashioned by the counter-actions of those threatened with punishment, who could take steps to reduce their risks. Workers and managers diverted effort from production into mutual insurance: since the threat was shared among them, they could agree to cover up each other’s shortcomings. At the grass roots, Berliner (1952) showed how managers, accountants, and technical workers in enterprises covered up poor results. Expressions in everyday use like “pulling wool over [the authorities’] eyes” [ochkovtiratel’stvo], and pripiski and tufta, the civilian and Gulag slang
for exaggerated or non-existent output, encapsulated the constant effort to mislead authority. The necessary collusion relied on horizontal trust accumulated through “acquaintance and relationships” [znakomstvo i sviazi]. The archives have added detail: Filtzer (2002) has shown that postwar managers tolerated lateness and absence to maintain goodwill, and colluded with workers against the centre to underreport violations. Managers did, however, pursue illegal quitters who undermined morale and the factory’s capacity to fulfil the plan. It was the rural police and courts that then gave moral shelter to the young offenders who had run home. In all these ways, mutual insurance tended to cut the individual risk of punishment. The archives also show how collusion travelled vertically upward through the hierarchy. Harris (1999, pp. 156-63) has described how the Urals regional leadership formed itself into a “protected, mutually reliant” clique that fended off criticism from below and investigation from above in the mid-1930s, but was eventually wiped out in the Great Terror. Belova (2001) has shown that high-level patrons could protect the most egregious embezzlers.

Faced with widespread enforcement failures at lower levels, officials responded in two ways. First, they aimed to undermine horizontal trust by rewarding whistle blowing and informing (Belova 2001). Second, Stalin forced the legal system, local party offices, and the militia to increase arrest and conviction rates or suffer penalties themselves. The most common method of forcing repression was to distribute quotas by region and profession to officials at lower levels (Danielson et al. 2005, vol. 1). In the Great Terror of 1937/38 local officials had to work feverishly to achieve a set number of confessions per day (Vatlin 2004). To fulfil such plans the police officials imputed individual guilt from increasingly trivial differences in behavior. Whether or not these measures reduced the Type II errors, they seem likely to have encouraged false denunciation and confession and so added to the errors of Type I.

Team Stalin hoped, we suppose, that increased coercion would induce agents to take a lower wage without withdrawing effort, making coercion effectively costless. The evidence suggests that in practice those threatened with punishment raised the chance of a Type II error through mutual insurance, but this diverted effort from production. At the same time successful insurance against Type II errors combined with the increased chance of a Type I error to blunt the effect of penalization on the reservation wage. Finally, the heavy direct costs of the repressive apparatus must have further reduced the net payoff from coercion.

At the moment we have only one empirical study of the aggregate effects of the terror on industrial production (Katz 1975), which attributes the 1937 slowdown in industrial growth to adverse effects of political repression. New case studies show the disastrous effects of the Great Terror on important Gulag operations (Ertz 2003). Finally, the archives have shown that Soviet leaders themselves saw a direct connection between the terror and falling productivity (Davies 2004): the terror eliminated a generation of skilled and knowledgeable managers and technical staff, worsened labor discipline, and damaged effort.
4. Conclusion

The study of Stalin’s economic decision making suggests how we can understand the motivation behind his application of coercion to the labor force: he expected it to increase the surplus of resources at his disposal. Closer analysis shows, however, that coercion did not yield the results expected. Harsh punishments for violations did not create obedient citizens or a law-abiding society. People who were threatened or frightened did not just do as they were told. Instead, they allocated resources to protecting themselves in ways that detracted from Stalin’s objectives.

It is sometimes suggested that the Soviet economic system could not have been held together without coercion. Whether or not this is true in some ultimate sense, it seems clear to us that the level of coercion in the Stalinist economy up to the early 1950s was counterproductive. In fact, those who managed the instruments of coercion were fully aware of it at the time: when Stalin died, they could hardly wait to dismantle important components of the coercive system.

References


