

Original citation:

Miller, Marcus, 1941- and Smith, Jennifer C. (2015) In the shadow of the Gulag : worker discipline under Stalin. *Journal of Comparative Economics*. ISSN 0147-5967 (In Press)

Permanent WRAP url:

<http://wrap.warwick.ac.uk/66181>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

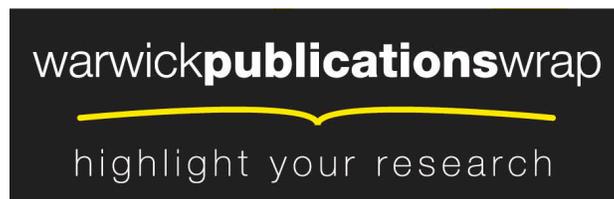
NOTICE: this is the author's version of a work that was accepted for publication in *Journal of Comparative Economics*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Journal of Comparative Economics*

<http://dx.doi.org/10.1016/j.jce.2015.01.005>

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



<http://go.warwick.ac.uk/lib-publications>

Accepted Manuscript

In the shadow of the Gulag: worker discipline under Stalin

Marcus Miller, Jennifer C. Smith

PII: S0147-5967(15)00020-7

DOI: <http://dx.doi.org/10.1016/j.jce.2015.01.005>

Reference: YJCEC 2437

To appear in: *Journal of Comparative Economics*

Received Date: 25 February 2014

Revised Date: 12 January 2015



Please cite this article as: Miller, M., Smith, J.C., In the shadow of the Gulag: worker discipline under Stalin, *Journal of Comparative Economics* (2015), doi: <http://dx.doi.org/10.1016/j.jce.2015.01.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

In the shadow of the Gulag: worker discipline under Stalin

Marcus Miller (corresponding author)

University of Warwick
Department of Economics
Coventry
CV4 7AL
UK
marcus.miller@warwick.ac.uk

Jennifer C. Smith

University of Warwick

ACCEPTED MANUSCRIPT

In the shadow of the Gulag: worker discipline under Stalin

January, 2015

Abstract

An ‘efficiency wage’ model developed for Western economies is reinterpreted in the context of Stalin’s Russia, with imprisonment – not unemployment – acting as a ‘worker discipline device’. The threat of imprisonment allows the state to pay a lower wage outside the Gulag than otherwise, thereby raising the “surplus” left over for investment; this externality provides a reason for coercion over and above the direct productivity of those in custody.

Just how credible the threat of imprisonment was under Stalin is documented using archival data now available; but the enormous scale of random imprisonment involved is, we argue, attributable not to economic factors but to Stalin’s insecurity in the absence of a legitimate process for succession.

We develop a model of demand and supply for industrial labour in such a command economy. To get more resources for investment or war, the state depresses the level of real wages; to avoid incentive problems in the wider economy, the harshness of prison conditions can be intensified. This is the logic of coercion we analyse.

Keywords: Labour discipline · asymmetric information · efficiency wage · Soviet Union · Stalin

JEL Nos: D82 · P23 · P26 · P27

1. Introduction

In the 1930s, when Western economies were plagued by the mass unemployment of the Great Depression, Josef Stalin could claim to have found a cure: a command economy with ambitious five year plans to catch up with the West by rapid industrialisation. The state would maintain high levels of aggregate demand to ensure full employment. But what about supply? Here Stalin had to confront the issue of incentives:¹ how was he to motivate workers with low levels of skill, including – as a result of collectivisation – “millions pouring in from the countryside entirely lacking in training or experience of the rigour and rhythms of life in a factory or on a construction site” (Acton and Stableford, 2005: 315)?

The Russian economy had been growing before the revolution (Gregory, 1982); and it was recovering and returning to growth in the 1920s, before the Gulag² was established. Transition to the command economy posed a major challenge for the dictator, however (Markevich and Harrison, 2011). Ample labour was available following the transfer of population from the countryside to industrial towns, but how was Stalin to extract the resources needed for rapid industrial development?

One approach would be to *borrow* from the workers the resources needed for investment. But recourse to capital markets was evidently not an option for Stalin: ideologically, private property was viewed as theft, and the livelihoods of millions of kulaks were expropriated on that basis. A second approach, analysed by Sah and Stiglitz (1984), would be to engineer a *terms-of-trade transfer* of resources from agriculture – essentially by reducing the cost of food.³ A third approach – more consistent with the evidence according to Carter (1986) and Ellman (1975) – is to enforce a *wage reduction* in the industrial sector itself, so the workers in the newly industrialising state pay for the capital needed to accelerate economic growth: “The source of the increase in accumulation in 1929-32 was the surplus obtained from the employment of additional workers in the urban sector *at real wages less than those enjoyed by employed workers in 1928 plus the surplus obtained by reducing the real wages of those who had been employed in 1928*” (Ellman, 1975: 856, italics added).

This is the approach we explore in this paper. The basic insight is that the wage that has to be paid in “free” labour market depends on the conditions in the Gulag. The main channel through which the prison system contributed to Stalin’s surplus was this externality – its depressing effect on the efficiency wage in the non-Gulag economy. Our approach is thus differentiated from those that evaluate the Gulag as productive (or not) in its own right.

Historians in Russia have debated whether the Gulag was necessary to settle remote regions, extract timber and minerals, and build facilities; they have also discussed whether Gulag labour

¹ The incentive problems arising from asymmetric information are central to Stiglitz’s critique of the Soviet system in *Whither Socialism?* (1994).

² Gulag is the Russian acronym for the Chief Administration of Corrective Labour Camps and Colonies.

³ This calls for a two-sector analysis, with the choice of terms of trade between sectors known as the ‘scissors problem’.

was more or less productive (in gross or net terms) than “free” labour (e.g. Borodkin et al., 2013). In the framework of this paper, in contrast, whether the Gulag produced anything or nothing internally is not the key. Its main function was to produce an externality: to be a terrible place that ordinary workers wished to avoid, so they worked harder for less. An analogy may be useful to illustrate this type of externality – namely the Debtor’s Prison, a well-known and much-feared institution in Victorian England. As a mechanism for securing the payment of debts, it is surely the *ex-ante* incentive effect that matters: the productivity of the debtor in prison is largely irrelevant.

This analysis faces the same challenge as Shapiro and Stiglitz’s (1984) ‘efficiency wage’ account of the unemployment as a discipline device⁴ to prevent shirking: if the threat is fully credible, *there should be no shirkers in equilibrium*. How then to account for the observed unemployed? In the Western case, they argued, unemployment should be attributed not to shirking, but to on-going random job break-ups – the flux of capitalism. To account for the extent and persistence of the Gulag in the Russian case, we appeal to the politics of repression. In the absence of a legitimate process for assured succession, the Gulag served a key political objective: that of maintaining the dictator in power by random imprisonment for political reasons.⁵

The paper proceeds as follows. In Section 2, a stylised ‘efficiency wage’ model is used to consider how incentive constraints facing a dictator determined to catch up with the West may be addressed by the threat of incarceration in a labour camp. The reality of this threat is borne out by recently available data on the custodial population in Soviet Russia from 1917 to 1953, which is presented and discussed in Section 3.

In Sections 4 and 5 the Shapiro and Stiglitz approach is adapted to fit Soviet objectives and forms of coercion. Thus we assume that dictator aims to maximise the investible surplus, subject to an *efficiency wage* constraint on the supply of labour and a *survival* constraint on his hold on power. Random incarceration helps achieve survival, while harsh prison conditions help keep down the cost of civilian labour. This provides a stylised account of Stalin’s system;⁶ and the role of harsher punishment is re-examined in this framework in Section 6.

While it is not our objective to estimate an econometric model of the system, in Section 7 we use the Stalin era data now available to calibrate key accounting parameters determining the size of the Gulag; and we discuss informally how shifting views of the productivity of prison workers will affect this. We also discuss how, when the Gulag system was dismantled, the command system was exposed to existential challenges – including especially societal pressures imposed by the campaign for universal Human Rights. Section 8 concludes.

⁴ Where the ‘efficiency wage’ is the minimum that has to be paid to avoid ‘shirking’.

⁵ Kornai (1992) likewise argues that the collectivisation of agriculture, in Soviet Russia and elsewhere, also showed political imperatives being given precedence over economic objectives.

⁶ Skidelsky (1995), Wintrobe (1998) and Gregory et al. (2006) provide more developed political economy perspectives.

2. Efficiency wage theories

How was Stalin to motivate industrial workers in the new Soviet system? ‘Efficiency wage’ theories developed for Western economies may provide some insights. Akerlof and Yellen (1990), for example, emphasise how worker motivation depends on *fairness*: so workers’ effort depends positively on the wage (or the consumption level) they receive, up to the point where they are paid the ‘fair wage’ and supply ‘full’ labour effort. This approach is used by Gregory (2004) to explain the trade-offs involved in choosing between consumption and investment in a command economy. The dictator, wishing to maximise investment, will pick a wage lying below the ‘fair wage’, but above a ‘strike wage’ at which workers will withdraw their labour. Gregory discusses how Stalin realised that consumption had to be increased to counter declining productivity in the early 1930s: and how he attempted to manipulate the fair wage by promises of a brighter future.⁷

The efficiency wage theory of Shapiro and Stiglitz (1984), by contrast, focuses on *asymmetric information* and the need to prevent ‘shirking’. Assuming the supply of effort is all-or-nothing, the worker is paid to put in effort, but failure to do so (‘shirking’) will – if detected – lead to loss of employment and income. With full monitoring, wages will have to exceed unemployment benefit by enough to cover effort; but with imperfect monitoring an added premium is needed to increase the value of a job so as to deter unobserved shirking. These incentive problems, in conjunction with exogenous labour turnover, lead to the emergence of persistent unemployment as a ‘worker discipline device’ – an information-based rationale for ‘the reserve army of the unemployed’ discussed much earlier by Kalecki (1935).

As Wintrobe (1998: 28) remarks, however: “The prospect of unemployment is not the only possible discipline device”. The Soviet alternative considered here is being sent to a labour camp, where work is compulsory and conditions deliberately harsh.⁸ For if shirking is treated as a crime meriting imprisonment, efficiency wages will depend, *inter alia*, on prison conditions and on duration of punishment. To illustrate the ideas involved in this information-based approach, we start with the extreme case where shirking, if detected, would lead to incarceration without release, referred to as ‘dire punishment’ – a term used in repeated games to denote a state from which there is no transition. The reason for beginning with this admittedly extreme case is because it is straightforward to derive the ‘efficiency wage’ – the minimum wage needed to avoid shirking – and to appreciate the incentive effects of ‘carrots and sticks’, issues considered in more detail later.

2.1. Efficiency wages with ‘Dire Punishment’

⁷ Gregory (2004) also mentions the possible use of forced labour to incentivise workers – the principal focus of this paper.

⁸ Nalebuff and Stiglitz (1983) also consider the idea of ‘horrible punishment’, but this is in a context of heterogeneous labour where individual workers can reduce the probability that lack of effort is detected. Here the role of the Gulag is analysed as an incentive device for the average worker, assuming labour is homogenous and the detection probability is exogenous.

Let w denote the real wage and e denote effort while working, and assume (as in Shapiro and Stiglitz, 1984) that the welfare of one who works can be measured simply by $w - e$, the excess of wages over effort. Let q denote the hazard rate of being detected while ‘shirking’ (putting in zero effort) and assume those caught shirking are sent to prison, where the level of welfare is $\gamma = c_P - h$, i.e. the excess of prison consumption over the effort of hard labour. (Note that γ will be *negative* if the value of prison consumption fails to compensate for the hard labour inside prison – with implications taken up at the end of this section.)

When the efficiency wage is paid in the ‘dire punishment’ case, the effort saved by shirking is just matched by the prospect of losing one’s job and being sent to prison, which (in continuous time) can be written as

$$e = q(V_E - V_P),$$

where q denotes the likelihood of detection and the terms inside parentheses represent the capitalised values of working and imprisonment, $V_E = \frac{w - e}{r}$ and $V_P = \frac{\gamma}{r}$, r being the discount rate. Thus, for given q , there will be no shirking if the (capitalised) benefit of working *versus* being in prison, $(w - e - \gamma)/r$, is high enough.

On substituting and rearranging, one obtains the No Shirking Condition:

$$w^* = e + \gamma + re/q \tag{1}$$

which shows that the ‘efficiency wage’ w^* (just high enough to deter shirking) rises with the level of effort and the interest rate; and falls with the efficacy of monitoring and the harshness of prison conditions.⁹ Note that these effects on the supply-price of labour are due to a credible threat; and, if wages are at or above the efficiency level, there need be no shirking.

The basic insight of this formal analysis is to show how, without reference to production in the camps and colonies, the Gulag can nevertheless contribute to Stalin’s surplus via an externality: its depressing effect on the efficiency wage in the non-Gulag economy.

2.2. Dismissing the monitors

Crude as this characterisation may be, it can help explain why Stalin’s early attempt to dispense with monitors for ideological reasons was later abandoned. The First Five Year Plan was launched “with a wave of attacks on managers and specialists suspected of harbouring alien class sympathies” (Acton and Stableford, 2005: 316); but this turned out to be “incompatible with the discipline drive, given their direct involvement in monitoring labour performance and implementing measures designed to raise productivity”. The adverse incentive effects are obvious from equation (1): a fall in the probability of detection q raises the efficiency wage. As shown in Figure 1, a rise from w_A^* to w_B^* , will, for a given marginal

⁹ See Romer (2012: Chapter 10 Unemployment, p.474) for analogous results in the Shapiro/Stiglitz model..

productivity of labour, lead to a contraction of employment and a reduction of the resources available for investment – as measured by total output less the wage bill.

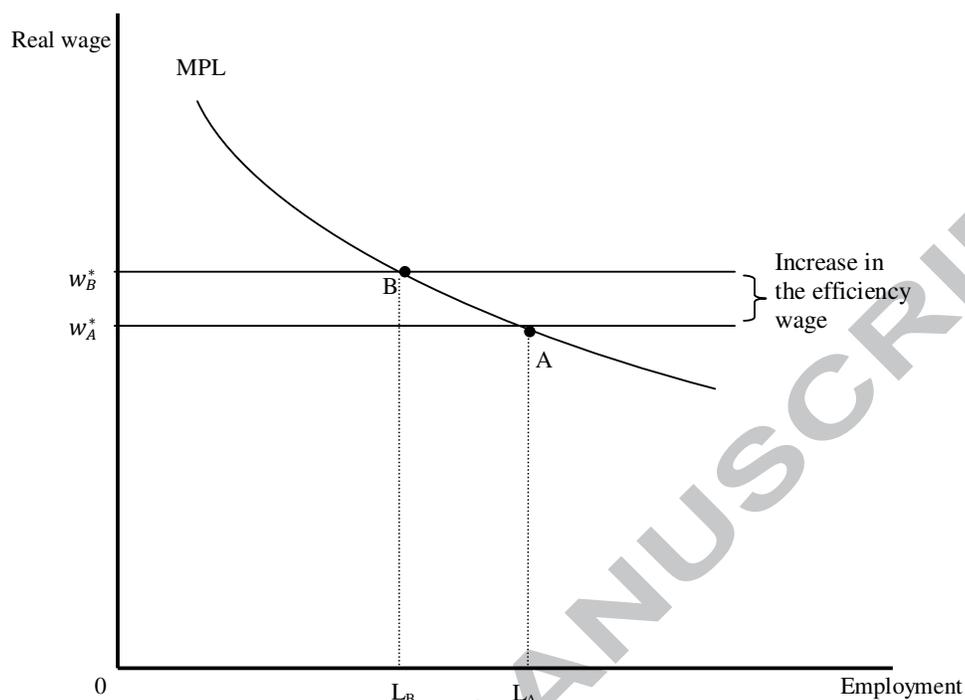


Fig. 1. Firing the monitors

By 1931-1932, however, Stalin had apparently learnt from experience, as the policy was reversed. “Specialists trained under the old regime, he announced, had seen the light and could now be trusted ... the authority, status and privileges of the white-collar strata now began to be energetically buttressed” (Acton and Stableford, 2005: 316). Before turning to the data and the inclusion of political factors, we use this simple framework to discuss two other facets of the Five Year Plans: the *carrot* of wage enhancement promised for the future, and the *stick* of harsher threats in the present.

2.3. The carrot: promises of a productivity bonus

In the First and Second Five Year Plans, Stalin argued that workers should accept restraints on their current wage in return for the promise that – thanks to higher investment – the supply of consumer goods would at least double, or perhaps even triple, by the late 1930s, see Davies (1984). Can the efficiency wage effectively be restrained by such “visions of a brighter future” (Gregory, 2003: 97)? Insofar as a future productivity bonus is like a financial asset that comes with the job, the answer is in principle positive: enhancing the value of the job by future bonuses should help deter shirking, so the level of the current ‘efficiency wage’ will fall. Stalin’s promises of a brighter future could, it seems, help solve the start-up problem by inducing a voluntary reduction in the current industrial wage.

But it is also possible – and incentive-compatible for the dictator – to pay no bonuses when the productivity boost arrives, but to increase employment¹⁰ instead: so the benefits of the productivity increase will go to the state and not to the existing workers. If this is how productivity improvements are in fact going to be handled, existing employees may be fooled early on; but, in time, the hoped-for effect on the efficiency wages will vanish as they realise what is in store.

Pure exhortation was also tried in Russia. Many managers hoped that others would follow the celebrated example of Stakhanov, who in 1935 mined far more coal per hour than the norm.¹¹ The formula for the efficiency wage confirms, however, that more effort calls for a higher efficiency wage; and “ordinary workers interpreted the Stakhanovite movement as a plot to extract more work for the same wage” (Gregory, 2004: 105).

2.4. *The stick: harsher punishment*

If promises and exhortation fail to provide adequate incentives, there is a brutal alternative open to a dictator: deprivation of liberty and manipulation of living conditions in prison. As Gregory and Harrison (2005: 740) put it: “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”.

Shapiro and Stiglitz (1984) show how the efficiency wage can be lowered by fiscal devices (such as labour subsidies funded by profit taxes). But, by violating the Human Rights of workers, coercion offers an alternative. Even without taking into account the loss of liberty that imprisonment involves, it is clear that the credible threat of coercion can promote the supply of resources to the state. This is clear from equation (1) where reducing the parameter $\gamma = c_p - h$, measuring the utility of prison consumption against the harshness of conditions, cuts the efficiency wage.¹² By the same token, amelioration of prison conditions will raise efficiency wages in civilian employment.

3. The Gulag

Instead of the reserve army of unemployed, the key instrument of coercion in Stalin’s system was, we contend, the prospect of imprisonment under harsh conditions. To judge whether imprisonment in the Gulag could act as a credible threat, we chart the size of the Gulag, how this changed over time, and the reasons for imprisonment.

¹⁰ With increased transfers of labour from agriculture, for example.

¹¹ He cut 102 tons of coal in 5 hours 45 minutes, beating the ‘norm’ of 7 tons by a factor of over 14 (Gregory, 2004: 103).

¹² Note that, with meagre rations and harsh working conditions in prison, the parameter γ will be negative, as will be its capitalised value, V_p . By the no shirking condition, the value of a job V_E need only exceed V_p by a constant: so harsh prison conditions could lower wages in civilian employment even below the cost of effort supplied.

3.1. The extent of the Gulag Archipelago¹³

Fig. 2 provides an overview of the Soviet prison population over the years 1917 to 1953, where we use the term ‘prison’ to encompass the whole of the Gulag system (generally understood to include prisons, colonies and camps).¹⁴

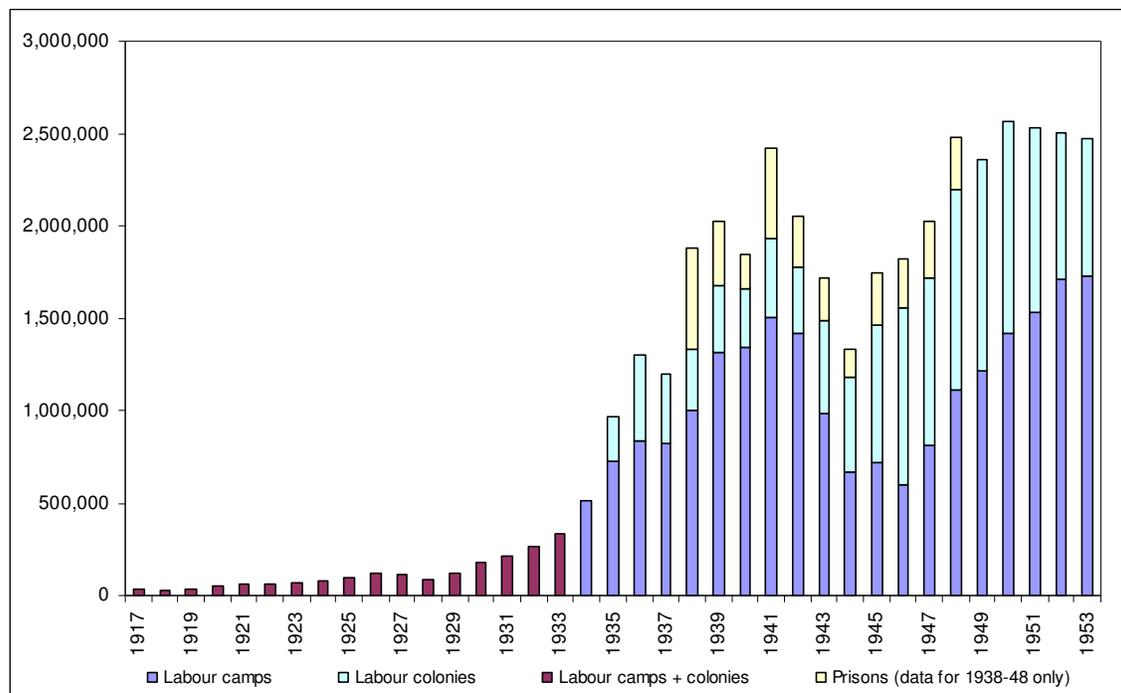


Fig. 2. USSR custodial population, 1917-1953

After Stalin and his allies took control of the Politburo in 1928-9, and the decision was taken to forcibly collectivise the peasants,¹⁵ numbers in custody began to rise sharply. The Law of Corrective Labour Camps of 1930 placed all camps and colonies into the Gulag system, and the harsher sentencing after 1930 brought small-time crooks into prison so that by 1934, when the NKVD¹⁶ took charge of the camp system, around half a million were in custody (Overy, 2004). As the NKVD tightened security and supervision, the possibility of escape diminished, and the number imprisoned more than doubled in a couple of years.¹⁷

¹³ As Solzhenitsyn (1974) described the prison system.

¹⁴ Labour settlements are excluded from the figures. Further details and sources are provided in Appendix 2.

¹⁵ While the focus of this paper is not on Stalin's agricultural policy, it is worth noting the scale of forced collectivisation: approximately 2.3 million kulaks were sent for resettlement between 1929 and 1940, most of them – 1.8 million – between 1929 and 1931 (Ivnitskii, 2004: 43; Zemskov, 1990).

¹⁶ People's Commissariat of Internal Affairs – the secret police.

¹⁷ As noted above, imprisonment was not the only discipline device open to the Soviet courts: the Coasian alternative of punishment via financial penalty was also used. What was termed ‘corrective work’ was quite common throughout the 1930s, constituting nearly half of all court sentences in 1935, for example (Getty et al. 1993: 1020). Typically, offenders were condemned to up to one year's ‘corrective labour’, the penalty consisting of work typically at the usual place of employment, with a reduction in the wage of up to 25 per cent and loss of credit for this service towards the length of service that gave rights to non-wage benefits such as vacation or pension (Getty et al. 1993: 1020; Sokolov, 2003: 32). The several laws on labour discipline passed in the late 1930s and early 1940s increased the numbers given non-custodial sentences – 1.7 million were given non-

The first peak in prison population is related to the Great Terror of 1937-1938. Initially, many perceived opponents were simply shot – casualties of Stalin’s political persecution¹⁸ that do not appear in the prison data. From 1938, however, imprisonment in labour camps became a more frequent punishment for political and economic crimes, driving the custodial population to a new peak of almost two and a half million. Due largely to releases connected with the war effort, however, the numbers in custody fell to a million and a third by 1944 (almost a million detainees were released to military service, often to the ‘storm’ units which suffered the heaviest casualties). For those left in the Gulag during the war, moreover, the mortality rate was extremely high: from 1941 to 1945, 1,005,000 inmates died in camps and colonies (Khlevnyuk, 2003: 51), due to scarce rations and the fact that the most able-bodied had been sent to war. After the war was over the custodial population rapidly resumed its upward march, reaching a plateau of two and a half million in 1948 (staying at this level until Stalin’s death in 1953, when more than half of all detainees were released).

3.2. *Reasons for imprisonment*

The Gulag held those judged guilty of political and economic crimes, as well as petty criminals. For present purposes, we highlight key aspects of Stalin’s approach to political opponents and how legislation was applied to punish perceived shirkers (see also Markevich, 2011 and Harrison, 2013).

Imprisonment was used primarily as an instrument of political power, with people being punished for ideological reasons. According to Lazarev (2003: 191): “The Gulag came into its own with the beginning of the Great Terror in 1937, when the upsurge in political prisoners drastically increased the population of the archipelago ... As the morose product of the tyrant’s paranoia, its main goal was to accommodate growing numbers of repressed opponents of the regime and ‘socially alien elements’ (like wealthy farmers and priests), while the economic use of prison labor was simply a by-product of the main political purpose”.

The aim of the Terror appears to have been “the removal of all strata of the population, which in the opinion of the country’s leaders were hostile or potentially hostile ... the Stalinist leadership always considered terror as its main method of struggle with a potential ‘fifth column’” (Khlevnyuk 1995: 167-169). Stalin’s suspicions encouraged pre-emptive action: “He did not only fear *hidden* enemies, who had not revealed themselves yet; he also feared *unconscious* enemies – those who did not even know themselves yet, who believed themselves to be loyal” (Harrison, 2008: 8; emphasis in original). “It was possible, evidently, to be a wrecker without meaning to be one or even knowing it. It was possible to wear a mask that

custodial sentences in 1940 (Getty et al, 1993: 1020) – but in relative terms the proportion sent to prison rose to more than two thirds in 1940 (compared to 20 per cent in 1930, for example; Solomon, 1980: 216).

¹⁸ “Those shot – largely adult males, many of them experts in various fields – were obviously not the targets of rational economic calculation.” (Robert Conquest, introducing Volume 1 of Kozlov et al. 2004: 30).

deceived even oneself” (Fitzpatrick 1999: 192; from a contemporary private diarist). (Harrison (2008) suggests that external threats may have had a role to play.¹⁹⁾

The ‘mass operations’ of the Great Terror lasted from July 1937 until November 1938.²⁰ Estimates vary, but even (conservative) archive data show that, from a working population of 66 million,²¹ 1.6 million (nearly 2.5%) had been arrested for counter-revolutionary violations by 1 November 1938 (Kozlov et al. 2004). As the conviction rate also rose – from around one third in 1930 to 85 per cent in 1937 (Gregory et al. 2006: 19) – the result was a huge rise in admissions to the prison system. There was no countervailing rise in releases – indeed, releases fell during the Terror – so the Gulag camp population rose: by 21 per cent during 1937 and by 32 per cent the following year (Gregory et al. 2006: 19).²²

Stalin also used the administrative and legal systems to combat economic crimes and to increase labour discipline. “The concept of economic sabotage was built into the Soviet Criminal Code of 1926. Any act of negligence or obstruction in the process of producing and distributing goods was defined as a counter-revolutionary crime, with penalties ranging from one year in jail to execution by shooting. During the First Five Year Plan additional laws were passed against the production of shoddy goods, against malpractice in retail stores and against state theft, all of them carrying penalties of between five and ten years in a camp” (Overy, 2004: 435).

To cope with absenteeism, lateness, drunkenness and high job turnover, tougher administrative measures were introduced in 1938,²³ and, between 1939 and 1940, new laws turned absence from work, tardiness, drunkenness and hooliganism into crimes²⁴ punished by up to four

¹⁹ “The fact is that Stalin was already personally secure in 1937 and his regime faced no immediate threat of any magnitude at home ... Stalin’s calculations were based primarily on his growing sense of the threat of a future war.” (Harrison, 2008: 10).

²⁰ During 1935-1936, Stalin had targeted the political elite, the three Moscow Show Trials enabling him to get rid of political rivals, who were forced to plead guilty to impossible charges of treason (these included Zinoviev and Kamenev, 1936; Piatakov, 1937; Bukharin, Tomsy and Rykov 1938) (Harrison, 2008: 7). In various communications and decrees of July 1937 – most notoriously, NKVD Order 00477 “About repression of former kulaks, criminals, and other anti-Soviet elements” – Stalin formulated plans for a terror campaign initially planned to start on August 5 and to last four months. Initial ‘limits’ for arrests and executions and the duration of the campaign had to be rapidly revised upwards to meet requests by local officials – many of whom were later punished for exceeding targets.” (Gregory et al., 2011).

²¹ In 1937.

²² Subsequent ‘national operations’ could also be seen as an overreaction to perceived threats. On January 31 1938 the Politburo issued instructions to the NKVD to complete by April 15 the destruction of the “counterrevolutionary nationalist contingent – Poles, Letts, Germans, Estonians, Finns, Greeks, Iranians, Harbinty, Chinese, and Romanians”, Bulgarians and Romanians (Khlevnyuk, 1995: 164; Harrison, 2008: 9-10).

²³ “On December 20, 1938, the Council of People’s Commissars (the highest state body) approved the decree ‘On the obligatory introduction of work books in all enterprises and institutions,’ a law designed to attack labor turnover and to reduce the free movement of labor among enterprises. Labor contracts were increased to five-year terms; all job changes, salary and reward histories, punishments, rebukes, and reasons for firings were registered in the labor book, which the cadres department used to evaluate workers’ performance” (Sokolov, 2003: 25).

²⁴ “In January of 1939, the Council of People’s Commissars decreed that tardiness of 20 minutes or more constituted an unauthorized absence from work. On June 26, 1940, the Presidium of the Supreme Soviet approved the decree ‘On the transition to an eight hour work day, a seven day work week, and the prohibition of voluntary departures of workers from enterprises and institutions’. The June 1940 law tied the worker to the enterprise and introduced criminal punishments for laziness, poor discipline, and tardiness. In August of 1940, criminal punishments were introduced for minor workplace infractions, such as drunkenness, hooliganism, and petty theft.

months in jail (Solomon, 1980: 217). These draconian new measures affected huge numbers: almost 1.8 million workers were convicted of absenteeism or lateness of more than 20 minutes in 1940 – more than two thirds of all criminal convictions (Solomon, 1996: 299); and there were over 4.5 million convictions during 1940-1942 (Sokolov, 2003, Table 2.1, p.28).²⁵ The use of labour laws clearly extended far beyond checking shirking; the random application of punishment effectively continued the terror of political persecution.

Pre-war labour discipline laws were retained after the war. Almost a quarter of a million workers were sentenced to jail terms for unauthorised absence, laziness or idleness in 1949; and of the 2.5 million imprisoned in the Gulag in 1950, half had been sentenced under the June 1940 labour discipline law.²⁶ Labour restrictions were eventually reduced in 1951,²⁷ but they were only fully abolished in 1956, after Stalin's death.²⁸

According to the model outlined in this paper, the dictator will choose to set wages high enough to eliminate shirking; and he can also worsen prison conditions (or increase monitoring) so as to reduce the efficiency wage that needs to be paid. But if shirking was successfully deterred in this fashion, what could account for the numbers in prison? The same issue arises in the Shapiro and Stiglitz model if, as they assume, unemployment is successful as a worker discipline device. As indicated in Section 1, their explanation for observed unemployment runs in terms of exogenous job break-ups (which occur in proportion the labour force) and of accessions (which are proportional to unemployment). Our account for the Gulag population is based on Stalin's sense of insecurity (sometimes called paranoia) leading to the brutal persecution of both those feared to be political subversives and vast numbers of random victims of laws purportedly designed to deter shirking. So the answer to the question posed is that the camps were full of political prisoners subjected to random terror and those imprisoned through the inappropriate application of labour laws.

4. Random imprisonment and the supply price of labour

If prison acts as an effective incentive device, those in prison will not be shirkers: as well as criminals and the victims of political repression, however, the prison population may include so-called shirkers caught by the random application of labour laws. The latter would include those caught by the 'five per cent rule' for denunciation, which Stalin expressed as follows:

The October 1940 reforms of vocational education raised the term of obligatory work after graduation to four years and prohibited voluntary departures" (Sokolov, 2003: 25).

²⁵ Not all of these served prison sentences: according to Sokolov (2003: 28), there were 955,000 prison sentences related to idleness and unauthorised departures during 1940-1941 (many more were sentenced to corrective work – see above). But the effect on the population was even bigger than these figures suggest: during 1940-1941 there were a total of 5.9 million trials for such offences (Sokolov, 2003: 27-28), which represents 2.7 per cent and 5.3 per cent of the workforce, respectively.

²⁶ The relative severity of punishment for these offences rose, as the numbers fined fell by half (Sokolov, 2003: 38 and 41).

²⁷ By a decree of the Praesidium of the Supreme Soviet of July 14 1951 'About the replacement of judicial responsibility of workers and employees for idleness, except in the case of multiple and extended absences with disciplinary and social actions' (Sokolov, 2003: 38).

²⁸ There was a steep rise in convictions to the Gulag for petty theft and similar offenses following a June 4, 1947 law (Kozlov et al., 2004: 49).

Your task is to check people at work and if something is not right, you must report it. Every member of the party, honest non-party members, citizen of the Soviet Union not only has the right but is obligated to report the deficiencies he sees. If they are right, maybe only 5 percent of the time, this is nevertheless bread. (Gregory et al., 2006: 18).

This edict seems designed to increase terror rather than promote detection,²⁹ particularly as there was “an official understanding during the 1937-1938 mass operations that a large number of innocent parties were to be convicted” (Gregory et al., 2011: 40). During the Great Terror and afterwards, indeed, “such a wide range of behaviors was criminalized that virtually every worker became liable to prosecution for something” (Gregory and Harrison, 2005: 739).

For workers faced with increased uncertainty concerning their liberty, random denunciation (and consequent punishment) at work will have an effect that is the reverse the promises discussed earlier: it will reduce the value of a job and increase the efficiency wage. Our analysis thus provides logical support for believing such random punishment was used for political rather than economic purposes: namely that it *raises* the efficiency wage and so *reduces* the investible surplus available.

To show this formally, assume the average worker treats the risk of imprisonment as a hazard that arrives randomly at rate πdt independent of effort supplied. The formula for valuing a job becomes

$$rV_E = w - e + \pi(V_P - V_E) \quad (5)$$

where the parameter π should be interpreted to include not just political repression but also the risk of being caught by false denunciation under the ‘five per cent rule’ discussed above and by other random applications of labour law.³⁰

The No Shirking Condition will be as before

$$q(V_E - V_U) = e \quad (6)$$

Assuming that people are liable to be sentenced to prison for a fixed term of length s with no expectation of early release,³¹ the valuation of ‘punishment’ becomes:

$$V_P = (1 - \exp(-rs)) \frac{\gamma}{r} + \exp(-rs)V_E \quad (7)$$

²⁹ There was little opportunity or effort made to stem opportunistic voluntary denunciations; denunciations made under torture were unreliable, often naming friends or acquaintances. Incentives for officials also promoted opportunism: “the NKVD itself opportunistically selected victims with large apartments that became a part of the NKVD inventory” (Gregory et al. 2011: 40, citing Vatlin, 2004).

³⁰ As Gregory and Harrison (2005: 739) note, “A broken-down commuter bus could make criminals of scores of hapless workers. 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 failed to report offenses 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”.

³¹ This formulation could be termed FIFO – first in, first out (equivalently, LILO) – and contrasts with Shapiro and Stiglitz (1984) where the accession to new jobs is random. For extreme values of the imprisonment rate (1 and zero), however, the efficiency wage is identical in both versions.

i.e. the weighted average of imprisonment and a job without imprisonment, where the weights depend on length of sentence, see Romer (2012: 508).

From (5), (6) and (7) we find

$$w^* = e + \gamma + \left(\frac{r}{1 - \exp(-rs)} + \pi \right) \frac{e}{q} \quad (8)$$

where s denotes sentence length. To a reasonable approximation when rs is small,³² this can be written

$$w^* = e + \gamma + (\rho + \pi)e/q \quad (9)$$

So the efficiency wage increases with the sum of the rate of incarceration and the release rate, where the latter is measured as $\rho = 1/s$.

In equilibrium where wages satisfy this condition, no-one shirks. What of the prison population? The dynamics of prison population and its steady state will, of course, depend also on the release rate (and on deaths in prison, which occurred on average at about one sixth of release rate³³). Treating the release rate as endogenous, however – as required to maintain the prison population constant in a steady state – one finds the efficiency wage depends on the prison population.

As releases must match incarcerations in the steady state, i.e. $\rho p = (1 - p)\pi$, we may summarize as follows:

Proposition 1 *There is a multiplicity of coercive equilibria in which the efficiency wage moves inversely to the proportion of the population in prison, p . Specifically the No Shirking Condition (NSC) is $w^* = e + \gamma + (\pi/p)e/q$ where π represents the rate of incarceration, to be accounted for in political terms.*

A detailed derivation is provided in Appendix 1, where it is shown that the efficiency wage is generally lower than that it would be for stochastic return to the labour force as in Shapiro and Stiglitz (1984). (This is because, in the stochastic case, for the same average release rate, individuals place greater weight on the relatively pleasant prospect of early release than on the unpleasant prospect of late release.)

³² Archive data indicate that the average sentence increased from 3.5 years in 1935 to 4.6 years in 1940, after the Terror led to a substantial increase in numbers being awarded long sentences for political offences (Getty et al. 1993). (Prior to October 1937, all sentences were for less than 10 years.) A combination of amnesties for those on short sentences and a sharp rise in average sentences for crimes such as theft during the 1940s led to further increases in average sentence, which reached 6.6 years by October 1945 (Alexopoulos, 2005).

³³ Data from Getty et al. (1993) reveal varying release rates: over the period 1934-1952, the release rate from camps (excluding colonies and prisons) varied between 15 and 47 per cent.

In Fig. 3(a), the efficiency wage is shown as the schedule labelled NSC for No Shirking Condition, sloping upwards from its intercept with the vertical axis (where the proportion of the population in prison, p , is unity (so $w^*(1) = e + \gamma + \pi e/q$).

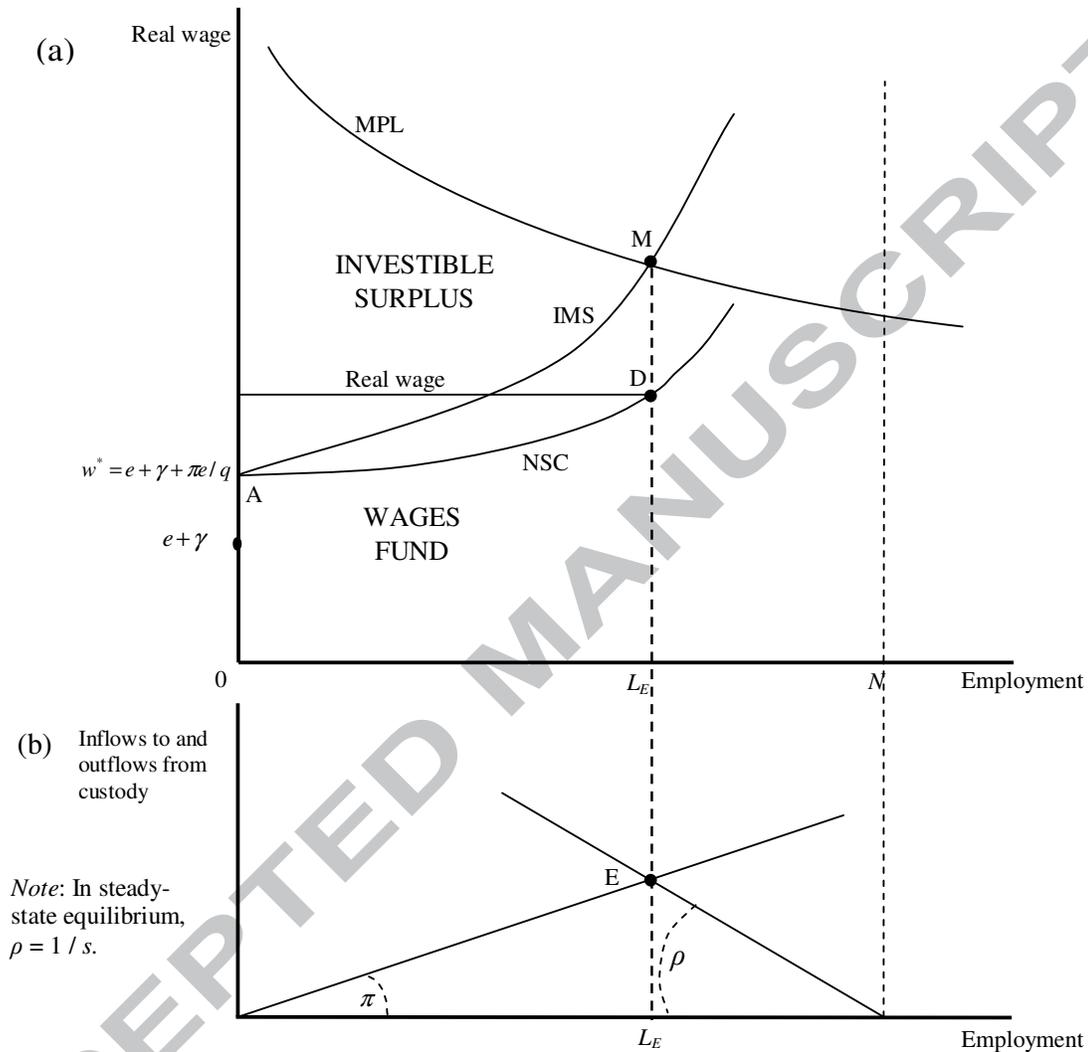


Fig. 3. Maximising resources for the state

5. Stalin's system: economic and political objectives and constraints

The No Shirking Condition, NSC, represents the effective supply price for labour. As for demand, we assume the dictator aims to maximise the surplus that can be extracted from the economy – subject to maintaining himself in power. While these resources can be used for investment or war or whatever the dictator chooses, we will, for expositional convenience,

assume that it is intended for investment – and we treat the maximisation of surplus first, before looking explicitly at the ‘survival constraint’.³⁴

5.1. Maximising the investible surplus

According to Skidelsky (1995: 102), “The Stalinist command economy, instituted in the First Five Year Plan in 1929, came to be seen as a bold solution to the problem of extracting enough saving from the population to pay for rapid industrialisation... The key technical requirement was limiting the wages fund and the supply of consumer goods, thus freeing up investment resources”. What does this involve? The answer we analyse is monopsony, i.e. we assume that:

Proposition 2 *The ‘technical requirement’ for limiting the resources used to fund employment, so as to maximise investible resources, is that the state act as monopsonist in the market for civilian labour.*

If L and w are chosen so as to maximize the surplus value available for the state consistent with the willingness of labour to work, treating the parameter π as predetermined implies that wages are forced down to their efficiency level and that the demand for labour restricted so that the efficiency wage lies below the marginal product of labour.

Specifically it is shown in Appendix 1 that state monopsony implies

$$F_L = w^* + Lw_L^* + v \quad (10)$$

where F_L indicates the marginal product of labour (shown as the curve MPL in the figure), w_L^* is the increase in the efficiency wage as civilian employment increases; and v is the ‘surplus value’, if any, extracted from the prison population (see Khlevnyuk, 2003, for discussion). Thus the MPL will exceed the NSC by two terms, the first measuring the impact of extra civilian employment on intra-marginal costs, the second being the opportunity cost of civilian employment in terms of prisoner productivity. If the Gulag was designed to be simply self-sustaining and no more, then $v = 0$. For convenience, this is the ‘neutral’ assumption made in the algebra to follow. In Section 7, however, the incentive to increase the Gulag population because of the perceived productivity of slave labour is discussed.³⁵

Note that this optimising condition (10) can be expressed in terms of model parameters by differentiating the NSC with respect to p , and p with respect to L , to yield an Investment Maximising Schedule (IMS) which determines the level of civilian employment. From the definition of the NSC in Proposition 1 and the definition of the proportion of the population imprisoned, $p = (N - L) / N$, it follows that

$$Lw_L^* = L(\partial w / \partial p)(\partial p / \partial L) = Le\pi / (qp^2 N) = e\pi(1 - p) / (qp^2).$$

³⁴ The formal analysis in Appendix 1 treats both issues simultaneously.

³⁵ See Solomon (1980) for a discussion of the build-up to, and effects of, Stalin’s 1929 edict on this subject.

With $\nu = 0$, setting the marginal cost of labour (paid at the efficiency wage) to match the marginal product of labour implies

$$F_L = w^* + Lw_L^* = e + \gamma + (r + \pi / p) e / q + e\pi(1-p) / (qp^2) = e + \gamma + (r + \pi / p^2) e / q \quad (11)$$

So the marginal cost of labour, shown as the steeper curve labelled IMS in Fig. 3(a), takes the same form as the NSC (the average cost), except that the parameter π is divided by p^2 instead of p .

Taking account of demand and supply, the equilibrium of the coercive system can be represented as in Fig. 3(a) where the intersection of the Investment Maximising Schedule (IMS) with the Marginal Product of Labour (MPL) at point M determines the level of employment and production; but the exercise of market power keeps the real wage on the NSC at point D. Since total wage costs are determined as the product of the real wage at D and the level of civilian employment L_E , the remainder of output – the area above the real wage line – will be available as the Investible Surplus.

The steady-state nature of the equilibrium is indicated by the matching of inflows and outflows at point E in the lower panel (where inflows to custody are measured as the rate of incarceration *times* the civilian labour force, πL_E , and outflows are given by releases *times* the custodial population, ρP , where the custodial population is $P = N - L_E$, and the parameter ρ is treated as endogenous).

5.2. Political survival

Ronald Wintrobe (1998, Chapter 2) argues that while dictators may have enormous power, the lack of any agreed process for succession – and the consequent threat of assassination – induces insecurity. “The insecurity of the dictator results in excessive repression”, he observes, “Stalin is the classic modern example”. Janos Kornai (1992: 362); and others have gone further to suggest that the Communist command economy was not so much a device for securing efficient production, rather a means of cementing the power of the dictator.

To account for random terror (as represented by the parameter π , the rate of random incarceration), we appeal therefore to political factors – the dictator’s need to survive in particular.³⁶ To do this adequately would surely involve a much broader treatment – taking into account the collectivisation of agriculture, and the mass execution of political suspects, as in Kornai’s study, for example.³⁷ As a short-cut we simply assume that the dictator’s survival requires maintaining power above a certain level, and that power depends positively on repression.

Where repression is proxied by random imprisonment, this implies, *ceteris paribus*,

³⁶ “Stalin’s hyper-suspiciousness bordering on paranoia was fundamentally political” (Kotkin, 2014: 736).

³⁷ A chilling political rationale for the Great Terror of 1937-1938 – in terms of maintaining a supportive constituency by simply eliminating ‘enemies’ – has been developed by Gregory et al. (2006, 2011); and a political-economy account of the command regime is provided in Harrison (2002, 2008).

$$\pi \geq \pi_0$$

which is in effect the Dictator's Survival Constraint, where the minimum level of repression will presumably shift with perceived threats to the dictator's rule.

As is evident from Proposition 1, however, random terror is economically counter-productive for the dictator in the framework developed here as it increases the efficiency wage.³⁸ So a dictator seeking to maximise investible resources will satisfy the constraint with equality, setting $\pi = \pi_0$. Thus Proposition 2 may be rewritten as:

Proposition 2a *To maximise investment and stay in power, the dictator will act as monopsonist in the market for civilian labour, subject to the No Shirking Constraint on wages and the Dictator's Survival Constraint on the rate of imprisonment.*

Though the rate of imprisonment is now endogenous, equilibrium will remain monopsonistic, with π set at π_0 .³⁹ Acemoglu (2008) observes that inefficient institutions may be preferable for the politically powerful: the dictator's use of repression may be an example.

This then is Stalin's creation: a society where everyone works and substantial resources are generated either for investment or for military expenditure, whatever the dictator decides. No one is unemployed, but many are in prison as repression is politically expedient. The state commands a goodly share of national resources, but wages are pushed down to efficiency levels. Ironically, the outcome for labour is almost as if it faced a greedy capitalist who wanted to maximise profits and had the monopsony power to do so. More than that, however, the state as employer can manipulate the living and working conditions for those not in civilian employment too.

6. Economic incentives: promises and punishment

6.1. Promises of productivity bonuses

The idea that promising productivity bonuses in the future could induce employees to work for less today has already been discussed for the case of Dire Punishment. In that case, where the NSC is horizontal, it seemed only too likely that all productivity gains would be appropriated by the state: so the promise would ultimately lack credibility, only effective as long as it fools people. So what promises should people believe? In their investigation of transition from dictatorship to democracy Acemoglu and Robinson (2006: 14) argue that credible promises are those that are 'time-consistent': such a promise is not what the government says today, it is the present discounted value of what the government will have to concede tomorrow.

We could solve for 'time consistent' promises here; and how they will lower the NSC. But they would only factor in the possibility of a modest shift along the NSC curve, hardly the two- or

³⁸ It is shown in Appendix 1, however, that the value of a job V_E is independent of the parameter π ; so when π increases the efficiency wage rises just enough to offset the higher risk of losing one's job.

³⁹ This is, of course, due to the simplicity of the constraint and would change if survival depended on the size of the investible surplus as well, for example.

three-fold increase in consumption that was being talked of. It is possible nevertheless that, at least for the earlier Five Year Plans, people could have been fooled – so promises might have worked for a while.

6.2. Making the Gulag harsher

As noted previously in the context of Dire Punishment, a more direct way for a dictator – unconstrained by the need to respect Human Rights – to increase civilian work incentives is by making prison conditions harsher. Thus:

Proposition 3 *Reducing the standard of living in the Gulag will reduce the efficiency wage, and increase the supply of investible funds.*

This can be confirmed from Fig. 4, where increasing the harshness of life in the Gulag (reducing γ) lowers the intercept of NSC from A to A' and shifts labour market equilibrium from D to D'.

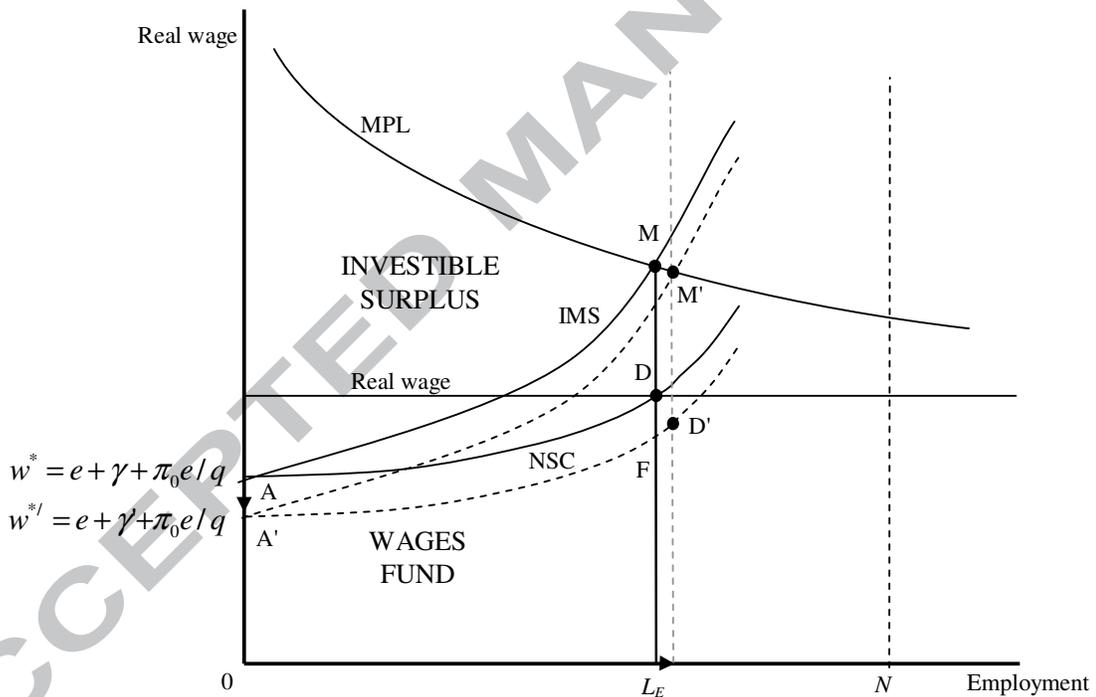


Fig. 4. The harsher the Gulag, the more resources for the state

The fall of the efficiency wage obviously increases investible funds at the pre-existing level of production, as measured by the distance DF times L_E . But the corresponding shift of the schedule IMS (which determines hiring) indicates that investible resources can be raised further by increasing civilian employment. With greater harshness of conditions in the Gulag, the dictator can pay less for labour while increasing output.

The logic of this section is as described by Gregory and Harrison (2005): to get more resources for investment or war, the actual wage must be compressed; and to avoid incentive problems the harshness of prison can be intensified. The end result is what Solzhenitsyn (1974) was to portray in *One Day in the Life of Ivan Denisovich*. Another characterisation of Stalin's creation is provided in Vasily Grossman's novel *Everything Flows*:

The most important principle of the State he constructed is that it is a State without freedom. ...in every sphere of life, freedom was overcome. It was the same with regard to factory work: work norms, pay, safety measures – all depended on the will of the State. From the Pacific Ocean to the Black Sea, non-freedom triumphed – everywhere and in everything. (Grossman, 2010: 207, 210)

In his *Economic Fables*, Rubinstein (2012) contrasts two mechanisms for allocating resources – the Market and the Jungle. The former is a competitive equilibrium with resources allocated by the Invisible Hand. In the latter might is right: so the lion eats first, then the tiger, etc. in an ordering determined by strength. By creating a command economy Stalin was of course seeking an alternative to the market: he promoted the role of a centralised powerful state led by himself – with the Gulag as icon.

It appears that the productivity of those in prison – in extracting timber and minerals for example – was an important element in planning the Gulag. But even when such plans proved unrealistic the system survived and grew – for at least two reasons. First and foremost, as discussed above, as an instrument for maintaining the dictator in power; second its role as an incentive device: like Dante's *Inferno*, the Gulag acted as a credible threat to workers who failed to fulfil the commands of the state. In the next section we illustrate the mechanics of the Gulag with partial equilibrium estimates of the pattern of inflows and outflows; and we discuss evolving views as to its role in Stalin's Soviet state.

7. Some data-fitting and some caveats

The archival data now available makes it possible to fit a basic model of the Gulag with quantitative estimates of prison flow rates over the Stalin era – including the impact of the drastic steps taken by Stalin to preserve his own power and to defeat Hitler. We discuss informally how shifting views of the productive potential of the Gulag could affect the prison population; and how these could be included in a more general model of Stalin's system.

The dynamics of the Gulag Archipelago: a flows perspective

In a partial-equilibrium context, inflows and outflows may be taken to bear a stable relationship to the associated population variables (of employees and prisoners respectively) with the exception of two periods: the Great Terror, when incarcerations increased at an alarming rate; and the Great War, when prison inmates were 'released' for military service.

Assuming specifically that inflows (imprisonments) are proportional to civilian employment L and outflows (reflecting releases, escapes and deaths in custody) are proportional to the prison population P , both measured in logarithms, the annual change in the prison population over

year t , ΔP_t , will be given by $\Delta P_t = \pi L_{t-1} - \rho^+ P_{t-1}$, where π is the inflow rate and ρ^+ denotes releases *plus* escapes and deaths.⁴⁰ Thus inflow and outflow rates can be estimated as coefficients in a regression of changes in prison population on lagged levels of employment and prison population with zero/one dummies for the periods 1937-1938 and 1941-1945 to allow for the effects of Terror and War; and an additional error term (assumed to be i.i.d. Normal) to capture random shocks to the system.⁴¹

Fitting this to the data over the Stalin era, 1928-1953, with interactive terms to allow for the higher incarceration rate during the Great Terror and the larger release rate during the Great War, gives the following coefficient estimates:⁴²

$$\Delta \hat{P}_t = 0.006L_{t-1} - 0.198P_{t-1} + 0.005L_{t-1}Terror_{t-1} - 0.101P_{t-1}War_{t-1} \quad R^2 = 0.30; N = 25$$

(0.002) (0.100) (0.004) (0.088)

The coefficient on L_{t-1} implies that, on average (excluding periods of Terror and War) about *half of one percent of the civilian workforce was incarcerated each year* while the coefficient on P_{t-1} implies that around one fifth of prisoners were released (or died in custody) – outflows which averaged around 400,000 annually.⁴³ The interaction terms imply, however, that the inflow rate *nearly doubles during the Terror*; and the outflow rate *increases by half during the War* (though these latter estimates are statistically not well determined).

Dividing through by labour force N (assumed to be constant and including those in prison, so $N = L + P$), we can write: $\Delta p = \pi - (\pi + \rho^+)p$, where $p = P/N$ and $\Delta p = \Delta P/N$. So in steady state where $\Delta p = 0$, the percentage of the population in prison will be $p = \pi / (\pi + \rho^+) = 0.006 / (0.006 + 0.198) = 0.029$ on the basis of the estimated coefficients. That is to say, using average values of the inflow and outflow coefficients and ignoring the shocks of Terror and War, *roughly three per cent of the working population would be in prison*. During Stalin's rule the working population (approximated by employment) averaged 67 million, so the coefficient 0.029 implies a steady-state prison population of 1.95 million. That this exceeds the .exceeds the simple average of the prison population over the period 1928-52 reflects the fact that the Gulag grew over time (see Fig. 2 and Appendix 2 Table A1).

The estimates above could be extended to take explicit account of the productivity of labour either in civilian production or in the Gulag itself. How would such demand-side factors affect our estimates? It will call for variation in the prison release rate: if the productivity of free

⁴⁰ Alexopoulos (forthcoming) emphasises that the data do not allow a reliable distinction between releases and deaths: Gulag managers faced incentives to release prisoners on the verge of death to hold down camp mortality statistics and because those individuals were no longer economically productive.

⁴¹ Annual data are used, so the absolute change in prison population between January 1 in year $t-1$ and January 1 in year t is regressed on the labour force and prison population measured on January 1 in year $t-1$. Note that the intercept is zero.

⁴² Standard errors are given in parentheses.

⁴³ Based on average workforce (approximated by employment) of just over 67 million and average prison population of nearly 1.6 million during 1928-52 (sources: Moorsteen and Powell, 1966: 643, Powell, 1968: 35, Luneev, 2006, Getty et al., 1993: 1048-49). That the coefficients imply that inflows exceeded outflows reflects the rise in the prison population that occurred over the period as a whole.

labour improves, for example, increased retention of workers in the civilian labour force will imply a higher release rate of prisoners. In a general equilibrium, release rates will also need to adjust in response to changes in the imprisonment rate π and to v , labour productivity within the Gulag. As equation (10) indicates, v places a wedge between the marginal productivity of labour and its marginal cost. In principle, changes in v will alter the opportunity cost of civilian labour, which would affect the number of people imprisoned by a resource-maximising dictator.

Contemporary views on the potential productivity of prison resources evolved over time. Initially the architects and managers of the Gulag from Stalin downward wanted it to be internally productive (Khlevnyuk, 2003). They did not value the Gulag only for its externality; they wanted to extract something from the Gulag itself. Consequently they had to manage the Gulag's own no-shirking constraint which, as time passed, induced experiments with labour incentives including wage bonuses and early release schemes (Borodkin and Ertz, 2003, 2005; Ertz, 2005). By the end of Stalin's rule, the Gulag's managers were increasingly convinced that it had become a financial drain on the state (generating no internal surplus). Not only that, but there were negative externalities to be considered too, as those released were spreading Gulag criminal culture through Soviet society. (Stalin did not agree, but as soon as he died they moved to close the Gulag down, Tikhonov, 2003.)

Thus contemporary views are consistent with a decline in the opportunity cost of civilian labour represented by productivity within the Gulag. This decline in the wedge is contrary to the expansion of the Gulag evident from Fig.2; so perceived internal Gulag productivity cannot be the primary explanation for Gulag dynamics. It is worth noting that our estimates have already taken account of a 'third use' of Soviet labour, namely to fight in the Great Patriotic War. Over the relevant period, this represented an opportunity cost to both free and Gulag labour, a factor we have tried to capture in our econometric account.

To estimate a model taking account of the full variation in demand side factors is left as a topic for future work. Continuing archival research is providing some of the necessary data, and disaggregation to the level of individual camp or region providing further evidence on internal productivity.⁴⁴

Challenges to the command economy

It appears that post-Stalin communism did for some time find a recipe for stability without terror and slave labour, but this was not conducive to economic efficiency. It should not be a surprise that a system that relies on depriving people of their human rights faces increasing problems as human capital becomes a more important factor in the production process.

The Soviet system after Stalin faced yet further challenges. Western spending on an ambitious and expensive arms race gave the Russian government added reason to compress wages – to produce arms as well as investment. But the progressive shifting of production to the service

⁴⁴ See, for example, Borodkin and Ertz, 2003, 2005; Ertz, 2005; Alexopoulos, 2005, forthcoming; and the disaggregated data on individual Gulag camps mapped by a UK ESRC-funded project, <http://www.gulagmaps.org/> and <http://www.memo.ru/>.

sector, where monitoring is more difficult, tended to *raise* efficiency wages, as Harrison (2002) points out. So too did the spreading doctrine of Human Rights, making repression ever less acceptable.⁴⁵ As Skidelsky (1995: 111) put it: “The command economy – the people – had to be flogged even harder. But neo-Stalinist incentives no longer worked”. From an efficiency wage perspective, these were existentialist pressures: pushing wages below efficiency levels means the command economy will collapse.

8. Conclusion

At a time when Western economies were prostrated by mass unemployment, Stalin’s command economy secured full employment. This owed little to Western ideas, however: Keynes had yet to write the *General Theory* (1936). Dictatorial command not only maintained aggregate demand, it was also used to solve the problem of incentives on the supply side; and here, we argue, the existence of labour camps played a key role.

The motivation for the frenzy of punishment in the late 1930s, and for the random application of coercive labour laws under Stalin was we believe primarily political. Nonetheless the system of labour camps generated a powerful externality which operated effectively as an alternative to unemployment as a labour discipline device. This we have shown by adapting the approach of Shapiro and Stiglitz to match Soviet conditions, an exercise that reveals how the supply price of labour for civilian employment can be lowered by making labour camp conditions harsher.

Coercion surely helped Russia to industrialise at high speed⁴⁶ and to produce the arms needed to defeat Hitler; but, as for the French Revolution, the regime of punishment became a monster – including mass executions of those perceived to be enemies of the state. Even when the Gulag system itself and some of the harsh labour laws were dismantled after Stalin died, the command system continued. But, as Sakharov and others testified, this involved maintaining the supremacy of the party over the rights of citizens.⁴⁷ An important issue for future research is whether an economic system whose operation rests upon the systematic violation of Human Rights is, in a globalised world, exposed to the risk of collapse – not from a shortage of demand, like the West in Depression, but from failure of supply, like an economic heart attack.

Acknowledgements

We are indebted to Mark Harrison and Andrei Markevich for expert advice on Soviet economic history and data; and for comments by Chris Doyle, Paul Gregory, Gregg Huff, Sharun Mukand, Paulo Santos-Monteiro, Joseph Stiglitz, Robert Skidelsky, Lei Zhang and by seminar participants at Sheffield, Warwick, the New School of Economics in Moscow and the European Association for Comparative Economic Studies. The paper has benefitted greatly from the detailed comments, suggestions and references of two anonymous referees.

⁴⁵ The Soviet Union abstained from endorsing the Universal Declaration of Human Rights when it first emerged in 1948. The ratification of the two Covenants on Civil and Political, and on Economic, Social and Cultural Rights in 1966, and their coming into force a decade later, are manifestations of the increasing Human Rights pressure.

⁴⁶ For the years of the Soviet Great Leap Forward, Ellman (1975, p. 844) points out that “measuring in 1928 Soviet prices, during the First Five Year Plan the volume of investment more than quadrupled”, rising from circa 15% of the national income to circa 44% in 1932.

⁴⁷ “Enshrined in Article 6 [of the new Constitution of 1977] and spelled out with greater clarity even than in the Stalin Constitution of 1936 was the central political role of the party.” (Acton and Stableford, 2007: 324).

Appendix 1. Derivation of Propositions

Proposition 1

The arbitrage condition for a job given random threat of imprisonment at the rate πdt is:

$$rV_E = w - e + \pi(V_P - V_E).$$

The No Shirking Condition, equalising the value of working and shirking, requires

$$V_E - V_P = e/q$$

as noted previously.

Let the valuation for punishment of a fixed sentence of length s be written as:

$$V_P = (1 - \exp(-rs))\frac{\gamma}{r} + \exp(-rs)V_E$$

i.e. the weighted average of imprisonment and a job, where the weights depend on length of sentence. (So $V_P \rightarrow \frac{\gamma}{r}$ as $s \rightarrow \infty$, the value used above for 'dire punishment'.)

As $\exp(-rs) \approx 1 - rs + (-rs)^2/2 + \dots$, so to a first order approximation $1 - \exp(-rs) = rs$.

The valuation function may be rewritten as

$$V_P = s\gamma + (1 - rs)V_E.$$

For given values of r , s and π , these three equations may be written as a linear system

$$\begin{bmatrix} 1 & -(\pi + r) & \pi \\ 0 & q & -q \\ 0 & -(1 - rs) & 1 \end{bmatrix} \begin{bmatrix} w \\ V_E \\ V_P \end{bmatrix} = \begin{bmatrix} e \\ e \\ s\gamma \end{bmatrix}$$

and solved as

$$\begin{bmatrix} w \\ V_E \\ V_P \end{bmatrix} = 1/\Delta \begin{bmatrix} qrs & r(1 + \pi s) & qr \\ 0 & 1 & q \\ 0 & (1 - rs) & q \end{bmatrix} \begin{bmatrix} e \\ e \\ s\gamma \end{bmatrix} \text{ where } \Delta = qrs.$$

For the efficiency wage in particular one obtains the result that

$$w = e + \gamma + \frac{e}{q} \left(\pi + \frac{1}{s} \right).$$

Note that the equilibrium values for V_E and V_P are independent of π . Hence from the arbitrage equation for the value of a job, it follows that the efficiency wage adjusts to offset any change in π .

Note also that this formulation for the efficiency wage can be cast in terms of imprisonment and release rates, namely

$$w = e + \gamma + \frac{e}{q}(\pi + \rho)$$

as, in steady state, the release rate is the reciprocal of sentence length, i.e. $\rho = \frac{1}{s}$.

In steady state equilibrium when outflows equal inflows with a sentence length of s , the prison population will be the number incarcerated per period, πL , multiplied by the sentence length, i.e. $P = \pi L s$. Hence $p = P/N = \pi(1-p)s$, where p denotes the percentage of the workforce in prison; and the release rate can be written as $\rho = \frac{1}{s} = \pi \frac{1-p}{p}$. This implies that

$$\pi + \rho = \pi + \pi \frac{1-p}{p} = \pi/p.$$

Given that $\rho = 1/s$, one may define a multiplicity of steady state equilibria. Substituting for $\pi + \rho$ in the equation defining the efficiency wage one obtains the result

$$w = e + \gamma + \frac{e}{q}(\pi/p)$$

i.e. the efficiency wage falls as the proportion of the labour force in prison increases.

Note finally that, in comparison with the analogous formulation in Shapiro and Stiglitz (1984), this definition omits a term involving the *outflow* from the state of punishment – in their case a/q , where a is the stochastic outflow rate from unemployment. This is *because the prison population must be higher in the stochastic-release model* (even when the average release rate is the same), due the fact that in stochastic models the discount rate leads individuals to place greater weight on the relatively pleasant prospect of early release than on the unpleasant prospect of late release. The level of the efficiency wage when the imprisonment rate is 1 and when it tends to zero are identical in both versions of the model.

Propositions 2, 3, 4 Maximisation of investment subject to labour incentives; political survival of the dictator; and survival in the Gulag.

For completeness, we analyse the case where the dictator is optimising subject to labour incentive, together with political *and* prisoner survival constraints.

The last of these reflects the fact that, for the dictator, the standard of living in the Gulag is a factor he can control. So prisoner survival is added as another constraint on the maximisation problem facing the dictator – adding the requirement that $\gamma \geq \gamma_0$.⁴⁸ The implication of maximisation subject to this constraint is clear enough: for the dictator to make conditions as

⁴⁸ When conditions became unbearable, prisoners would confess to capital crimes; see Acton and Stableford (2005, Documents 196 and 200) on the treatment of Bukharin and Meyerhold, for example.

harsh as possible conditional on prisoners' survival as well as his own, (so $\gamma \geq \gamma_0$ as well as $\pi \geq \pi_0$).⁴⁹

The dictator's problem can be written as

$$\max_{L, w, \pi, \gamma} f(F(L) - wL) \text{ subject to } w - w^*(L; \pi, \gamma) \geq 0, \pi \geq \pi_0 \text{ and } \gamma \geq \gamma_0$$

where w^* denotes the efficiency wage of Proposition 1, i.e.

$$w^*(L; \pi, \gamma) = e + \gamma + \frac{e}{q} (\pi / p) .$$

For this we form the Lagrangian

$$\Lambda(L, w; \pi, \gamma) = F(L) - wL + v(N - L) + \lambda(w^*(L, \pi) - w) + \lambda_1(\pi_0 - \pi) + \lambda_2(\gamma_0 - \gamma)$$

to be maximised subject to the Kuhn-Tucker conditions.

The First Order Conditions are

$$\partial \Lambda / \partial w = -L - \lambda = 0 \text{ so } \lambda = -L;$$

$$\partial \Lambda / \partial L = F_L - w - \lambda w_L^* - v = 0 \quad \text{where } w_L^* = w_p^* \frac{\partial p}{\partial L} = -w_p^* / N > 0;$$

$$\partial \Lambda / \partial \pi = \lambda w_\pi^* - \lambda_1 = -L w_\pi^* - \lambda_1 = 0 \quad \text{where } w_\pi^* = e / pq > 0;$$

$$\partial \Lambda / \partial \gamma = \lambda w_\gamma^* - \lambda_2 = -L w_\gamma^* - \lambda_2 = 0 \quad \text{where } w_\gamma^* = 1 > 0.$$

Since the objective is non-stationary in w , π and γ , and the Lagrange multipliers are negative, it is optimal to set $w = w^*(L)$, $\pi = \pi_0$ and $\gamma = \gamma_0$, the negative multipliers indicating that an increase in efficiency wages, in the repression level needed to keep the dictator in power and the conditions needed to ensure survival in the gulag will all reduce the optimal investible surplus.

With 'optimal' values for π and γ substituted into the problem, maximisation becomes simply:

$$\max_L F(L) - wL \text{ subject to } w - w^*(L; \pi_0, \gamma_0) \geq 0$$

$$\text{where } w^*(L; \pi_0, \gamma_0) = e + \gamma_0 + \frac{e}{q} (\pi_0 / p)$$

With the Lagrangean

⁴⁹ A more satisfactory approach would surely require explicit modelling of the conditions of life, work and incentives in the Gulag in some detail.

$$\Lambda(L, w; \pi_0, \gamma_0) = F(L) - wL + v(N - L) + \lambda(w^*(L; \pi_0, \gamma_0) - w)$$

the First Order Conditions are

$$\partial\Lambda/\partial w = -L - \lambda = 0$$

$$\partial\Lambda/\partial L = F_L - w - \lambda w^*_L - v = 0$$

Note that, given the incentive constraint on wages, the second condition becomes $F_L = w^* + Lw^*_L + v$ as discussed in the text.

It is not necessary to assume optimisation on the parameters π and γ in this way; in which case the FOCs will need to be evaluated at whatever values are taken as predetermined. (See, for example, Madden (1986) for further discussion of the ‘envelope theorem’ being used here.)

Appendix 2. Custodial Population Data: sources and methods

Year	Numbers incarcerated			Total custodial population	Release rate (%)	Admission rate (%)	Counter- revol- utionaries	Counter- revolutionaries as % of camp population
	Prisons	Labour colonies	Labour camps					
1917				34,083				
1918				26,888				
1919				33,948				
1920				47,863				
1921				62,544				
1922				60,559				
1923				71,545				
1924				77,784				
1925				92,947				
1926				122,665				
1927				111,202				
1928				85,158				
1929				118,179				
1930				179,000				
1931				212,000				
1932				268,700				
1933				334,300				
1934			510,307	510,307	28.9	0.9	135,190	26.5
1935		240,259	725,483	965,742	29.1	0.7	118,256	16.3
1936		457,088	839,406	1,296,494	44.0	0.8	105,849	12.6
1937		375,488	820,881	1,196,369	44.4	1.0	104,826	12.8
1938	548,417	336,786	996,367	1,333,153	28.1	1.2	185,324	18.6
1939	350,538	355,243	1,317,195	2,022,976	17.0	0.6	454,432	34.5
1940	190,266	315,584	1,344,408	1,850,258	23.6	0.8	444,999	33.1
1941	487,739	429,205	1,500,524	2,417,468	41.6	1.2	420,293	28.0
1942	277,992	360,447	1,415,596	2,054,035	36.0	1.0	407,988	28.8
1943	235,313	500,208	983,974	1,719,495	34.2	0.6	345,397	35.1
1944	155,213	516,225	663,594	1,335,032	22.9	0.5	268,861	40.5
1945	279,969	745,171	715,506	1,740,646	47.1	0.5	283,351	39.6
1946	261,500	956,224	600,897	1,818,621	19.3	0.6	333,833	55.6
1947	306,163	912,704	808,839	2,027,706	24.1	0.9	427,653	52.9
1948	275,850	1,091,478	1,108,057	2,475,385	23.6		416,156	37.6
1949		1,140,324	1,216,361	2,356,685	14.7		420,696	34.6
1950		1,145,051	1,416,300	2,561,351	15.3		578,912	40.9
1951		994,379	1,533,767	2,528,146	16.6		475,976	31.0
1952		793,312	1,711,202	2,504,514	19.3		480,766	28.1
1953		740,554	1,727,970	2,468,524	54.2		465,256	26.9
1959				948,000				

Table A1 USSR custodial population, prisoner flows, and political prisoners

USSR custodial population, 1917-1953

Sources: Total custodial population 1917-1934: Luneev (2006). Prisons, colonies and camps 1934-1953: data from the State Archives of the Russian Federation (GARF) published in Getty et al. (1993).

Notes: Data on the population of Soviet labour camps, colonies and prisons were made available during glasnost' from the Soviet Central State Archive. Total custodial population does not include those in labour settlements, as is usual in the literature. (From the point of view of labour discipline, settlements did not perform the same function as camps and colonies, in that the average non-agricultural worker faced no risk of being sent to a settlement.) Figures for the prison population relate to January 15 except for 1938, which refers to February 10. The 1938 prison figure is taken from a note to the Table in Appendix (a) of Getty et al. (1993). Figures for labour colony and camp populations refer to January 1. The 1938 "colonies" figure here subtracts 548,417 from the figure given in Getty et al. (1993), as the latter included those in prison. We note that the 1942 colonies figure is 1,000 lower than that previously given by similar sources (tabulated in Bacon, 1992); this also affects the total custodial population estimate for 1942. Many of these figures have been widely cited since; for example, Overy (2004). Prison was generally used only on a temporary basis: following an arrest, an individual would generally pass through prison for investigation and interrogation. More often than not, this led to a conviction. Most convicts were sent to camps or colonies to serve out their sentences (Getty et al. 1993: 1019).

Labour camps had existed under the Tsars. Under the new Bolshevik regime, in July 1918 a new system of approximately 300 camps was set up by the Cheka secret police (Overy, 2004) to house political offenders (although by the middle of 1919 the camps were receiving criminal as well as political convicts – Solomon, 1980: 200). Camps were initially intended to be economically self-sufficient, with prisoners working to pay for their own upkeep (but not on jobs for the state). The labour was hard – but could be refused by leftist political prisoners – and conditions were harsh. In addition to the camps, from 1919, the Commissariat of Justice ran a system of labour colonies for prisoners convicted of petty crimes with sentences of less than three years. Conditions in the colonies were less harsh, resembling open prisons; often prisoners worked alongside criminals sentenced to labour duty but not incarcerated. The end of the civil war in 1922 brought the merging of the administration of the camps and colonies. The Cheka (OGPU) retained a small network of camps, primarily in the north, to house political opponents. Numbers of prisoners in camps and colonies rose steadily, from around 30,000 in the early Bolshevik years to over 100,000 in 1926-7. Solomon (1980: 202) estimates that the (Solovki) camp detainees in 1927-28 accounted for between 10 and 15 per cent of the total camp and colony population.

The annual figures mask quite substantial fluctuations in inflow rates within years. Bacon (1992: 1077) cites the case of a particular year. As can be seen from Table A1, in January 1942 there were 1,776,043 incarcerated in camps and colonies,⁵⁰ a decline of more than 200,000 compared to the camp population of 1,929,729 recorded a year earlier in January 1941. But this decline hides a rise and subsequent fall during 1941: at the start of the Great Patriotic War on 22 June, the camp population was recorded as 2,300,000 – so during 1941 there was a rise of around 400,000 then a decline of more than half a million.

Release and admission rates, 1934-1947

Sources: Releases: Getty et al. (1993). Admissions: Bacon (1994). Employment: Moorsteen and Powell (1966), except 1941-4: Powell (1968).

Notes: Admission rate is admissions as a proportion of total employment (full-time equivalents). Release rate is releases as a proportion of the prison population as at 1 January in the relevant year. The particularly high release rates during 1941-1945 are in part explained by releases to the armed forces. Of the 1.956 million released during that time, Getty et al. (1993: 1040) state that 975,000 were released to military service (particularly to punitive or 'storm' units, which suffered the heaviest casualties). However, political prisoners were generally barred from release to the army (Getty et al. 1993).

Political prisoners (counter-revolutionaries) in labour camps, 1934-1953

Source: Getty *et al* (1993).

⁵⁰ This figure (taken from Getty et al. 1993) is 1,000 less than that given in Bacon (1992).

References

- Acemoglu, D., 2008. *Introduction to Modern Economic Growth*. Princeton Univ. Press, Princeton, NJ.
- Acemoglu, D., Robinson, J.A., 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge Univ. Press, Cambridge.
- Acton, E., Stableford, T., 2005. *The Soviet Union: a Documentary History Volume 1, 1917-1940*. Univ. of Exeter Press, Exeter.
- Acton, E., Stableford, T., 2007. *The Soviet Union: a documentary history Volume 2, 1939-1991*. Exeter, University of Exeter Press.
- Akerlof, G.A., Yellen, J.L., 1990. The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics* 105(2), 255-283.
- Alexopoulos, G., 2005. Amnesty: the revolving door of Stalin's gulag. *Slavic Review* 64(2), 274-306.
- Alexopoulos, G, forthcoming. *Health and Inhumanity in Stalin's Gulag*. Yale Univ. Press, New Haven, CT.
- Bacon, E., 1992. 'Glasnost' and the Gulag: new information on Soviet forced labour around World War II. *Soviet Studies* 44(6), 1069-1086.
- Bacon, E., 1994. *The Gulag at War: Stalin's Forced Labour System in the Light of the Archives*. Macmillan, London.
- Borodkin, L.I, Ertz, S. 2003. Coercion versus motivation: forced labor in Norilsk. In: *The Economics of Forced Labor: The Soviet Gulag*. Gregory, P.R., and Lazarev, V. (Eds.), Hoover Institution Press, Stanford, CA, pp. 75-103.
- Borodkin, L.I, Ertz, S., 2005. Forced labor and the need for motivation: wages and bonuses in the Stalinist camp system. *Comparative Economic Studies* 47(2), 418-436.
- Borodkin, L. I., Krasil'nikov, S.A., Khlevniuk, O.V. (Eds.), 2013. *Istoriia Stalinizma: Prinuditel'nyi trud v SSSR: ekonomika, politika, pamiat': materialy mezhdunarodnoi nauchnoi konferentsii*. Moskva, 28-29 oktiabria 2011 g. Rosspen, Moscow.
- Carter, M.R., 1986. The economics of price scissors: comment. *The American Economic Review* 76(5), 1192-1194.
- Davies, R.W. (Ed.), 1984. *Soviet Investment for Planned Industrialisation, 1929-1937: Policy and Practice*. Berkeley Slavic Specialities, Berkeley, CA.
- Easterly, W., Fischer, S., 1995. The Soviet economic decline. *The World Bank Economic Review* 9(3), 341-371.

- Ellman, M., 1975. Did the agricultural surplus provide the resources for the increase in investment in the USSR during the First Five Year Plan? *Economic Journal* 85(340), 844-863.
- Ertz, S., 2005. Trading effort for freedom: workday credits in the Stalinist camp system. *Comparative Economic Studies* 47(2), 476-491.
- Fitzpatrick, S., 1999. *Everyday Stalinism: Ordinary Life in Extraordinary Times*. Oxford Univ. Press, Oxford.
- Getty, J.A., Rittersporn, G.T., Zemskov, V.N., 1993. Victims of the Soviet penal system in the pre-war years: a first approach on the basis of archival evidence. *American Historical Review* 98(4), 1017-1049.
- Gregory, P.R. 1982. *Russian National Income, 1885-1913*. Cambridge Univ. Press, Cambridge.
- Gregory, P.R., 2004. *The Political Economy of Stalinism: Evidence from the Soviet Secret Archives*. Cambridge Univ. Press, Cambridge.
- Gregory, P.R., Harrison, M., 2005. Allocation under dictatorship: research in Stalin's archives. *Journal of Economic Literature* 43(3), 721-761.
- Gregory, P.R., Schroder P., Sonin, K., 2006. Dictators, repression and the median citizen: an "eliminations model" of Stalin's Terror (data from the NKVD Archives). CEPR Working Paper 6014, December.
- Gregory, P.R., Schroder P., Sonin, K., 2011. Rational dictators and the killing of innocents: data from Stalin's archives. *Journal of Comparative Economics* 39(1), 34-42.
- Grossman, V., 2010. *Everything Flows* (Trans. & Eds. R. Chandler & E. Chandler), Vintage Books, London.
- Harrison, M., 2002. Coercion, compliance, and the collapse of the Soviet command economy. *Economic History Review* 55(3), 397-433.
- Harrison, M. 2008. The dictator and defence. In: Harrison, M. (Ed.) *Guns and Rubles: the Defense Industry in the Stalinist State*. Yale Univ. Press, New Haven, CT, pp. 1-30.
- Harrison, M. 2013. Secrecy, fear and transaction costs: the business of Soviet forced labour in the early Cold War, *Europe-Asia Studies* 65(6), 1112-1135.
- Ivnitskii, N.A., 2004. *Sud'ba raskulachennykh v SSSR*. Sobranie, Moscow.
- Kalecki, M., 1935. A macrodynamic theory of business cycles. *Econometrica* 3(3), 327-344.
- Khlevnyuk, O.V., 1995. The objectives of the Great Terror, 1937-38. In: Cooper, J.M., Davies, R.W., Perrie, M., Rees, E.A. (Eds.), *Soviet History, 1917-1953: Essays in Honour of R.W. Davies*. St Martin's Press, New York, NY.
- Khlevnyuk, O.V., 2003. The economy of the OGPU, NKVD, and MVD of the USSR, 1930-1953: the scale, structure and trends of development. In: Gregory, P. R., Lazarev, P. R. (Eds.), *The Economics of Forced Labor: the Soviet Gulag*. Hoover Institution Press, Stanford, CA, pp. 43-66.

Kornai, J., 1992. *The Socialist System: the Political Economy of Communism*. Clarendon Press, Oxford.

Kotkin, S. (2014). *Stalin, Volume 1: Paradoxes of Power, 1878-1928*. Penguin Press, New York, NY.

Kozlov, V.P., Afanasiev, Y.N., Chubar'ian, A.O., Conquest, R., Danielson, E., Emmons, T., Gregory, P., Khlevnyuk, O.V., Khristoforov, V.S., Kozlov, V.A., Mironenko, S.V., Solzhenitsyn, A.I., Sorokin, A.K., Yakovlev, A.N., 2004. *The History of Stalin's Gulag, Late 1920s – Early 1950s: Collected Documents in 7 Volumes*. Rosspen, Moscow.

Lazarev, V., 2003. *Conclusions*. In Gregory, P. R., Lazarev, V. (Eds.), *The Economics of Forced Labor: the Soviet Gulag*. Hoover Institution, Stanford, CA.

Luneev, V., 2006. *Crime and punishment in Russia*, Demoscope Weekly, 0239. Available at <http://demoscope.ru/weekly/2006/0239/>. Demoscope Weekly, Moscow.

Madden, P., 1986. *Concavity and Optimization in Microeconomics*. Basil Blackwell, Oxford.

Markevich, A., 2011. *How much control is enough? Monitoring and enforcement under Stalin*, *Europe-Asia Studies* 63(8), 1449-1468.

Markevich, A., Harrison, M., 2011. *Great War, Civil War, and Recovery: Russia's National Income, 1913 to 1928*. *Journal of Economic History* 71(3), 672-703.

Moorsteen, R.H., Powell, R.P. 1966. *The Soviet Capital Stock, 1928-1962*. Homewood, Irwin, IL.

Nalebuff, B.J., Stiglitz, J.E. 1983. *Information, competition and markets*. *The American Economic Review* 73(2), 278-283.

Overy, R., 2004. *The Dictators: Hitler's Germany and Stalin's Russia*. Allen Lane, London.

Powell, R.P., 1968. *The Soviet capital stock and related series for the war years*. In: Becker, A.S., Moorsteen, R.H., Powell, R.P. (Eds.), *Two Supplements to R. H. Moorsteen and R.P. Powell, The Soviet Capital Stock, 1928–1962*. Yale Univ. Economic Growth Center, New Haven, CT.

Romer, D., 2012. *Advanced Macroeconomics (4th Edition)*. McGraw-Hill Irwin, New York, NY.

Rubinstein, A., 2012. *Economic Fables*. OpenBook Publishers, Cambridge.

Sah, R.K., Stiglitz, J.E., 1984. *The economics of price scissors*. *The American Economic Review* 74(1), 125-138.

Sakharov, A., 1975. *My Country and the World*. Collins and Harvill, London.

Shapiro, C., Stiglitz, J.E., 1984. *Equilibrium unemployment as a worker discipline device*. *The American Economic Review* 74(3), 433-444.

Skidelsky, R., 1995. *The Road from Serfdom*. Penguin Books, London.

Sokolov, A., 2003. Forced labor in Soviet industry: the end of the 1930s to the mid-1950s. In: Gregory, P. R, Lazarev, V. (Eds), *The Economics of Forced Labor: the Soviet Gulag*. Hoover Institution, Stanford, CA, pp. 23-42.

Solomon, P.H., 1980. Soviet penal policy, 1917-1934: a reinterpretation. *Slavic Review* 39(2), 195-217.

Solomon, P.H., 1996. *Soviet Criminal Justice under Stalin*. Cambridge Univ. Press, Cambridge.

Solzhenitsyn, A.I., 1974. *The Gulag Archipelago 1918-1956: An Experiment in Literary Investigation*. Harper and Row, New York, NY.

Stiglitz, J. E., 1994. *Whither Socialism?* MIT Press, Cambridge, MA.

Tikhonov, A., 2003. The end of the Gulag. In: Gregory, P.R., and Lazarev, V. (Eds.), *The Economics of Forced Labor: The Soviet Gulag*. Hoover Institution Press, Stanford, CA, pp. 67-73.

Vatlin, A., 2004. *Terror Raionnogo Masshtaba*. Rosspen, Moscow.

Wintrobe, R., 1998. *The Political Economy of Dictatorship*. Cambridge Univ. Press, New York, NY.

Zemskov, V.N., 1990. Spetsposelentsy (po dokumentam NKVD – MVD SSSR). *Sotsiologicheskiie issledovaniia*. 11, 3-17.

In the shadow of the Gulag: worker discipline under Stalin⁵¹**January 2015****Marcus Miller**

University of Warwick

Jennifer C. Smith

University of Warwick

Highlights

- The Gulag had incentive effects on Soviet labour akin to unemployment in the West
- Stalin's use of imprisonment for shirking-type behaviour and terror is documented
- An efficiency wage model shows how the threat of prison raises investible surplus
- It helps explain use of monitors, promises of a better future and harsh punishment
- But random incarceration, used for political purposes, raises the efficiency wage

⁵¹ M. Miller

Corresponding author, Department of Economics, University of Warwick and Centre for Economic Policy Research (CEPR),

Gibbet Hill Rd, CV4 7AL, Coventry

e-mail: marcus.miller@warwick.ac.uk

Phone: +44(0)2476523048

Fax: +44(0)2476523032

J.C. Smith

Department of Economics, University of Warwick, Gibbet Hill Rd, CV4 7AL, Coventry

e-mail: jennifer.smith@warwick.ac.uk

+44(0) 2476523469