

DISCUSSION PAPER SERIES

No. 6621

**PUNISHMENT WITHOUT CRIME?
PRISON AS A WORKER-DISCIPLINE
DEVICE**

Marcus Miller and Jennifer C Smith

INTERNATIONAL MACROECONOMICS



Centre for Economic Policy Research

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP6621.asp

PUNISHMENT WITHOUT CRIME? PRISON AS A WORKER-DISCIPLINE DEVICE

**Marcus Miller, University of Warwick and CEPR
Jennifer C Smith, University of Warwick**

Discussion Paper No. 6621
December 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Marcus Miller and Jennifer C Smith

ABSTRACT

Punishment Without Crime? Prison as a Worker-Discipline Device*

An 'efficiency wage' model developed for Western economies is reinterpreted for Soviet Russia assuming that it was the Gulag not unemployment that acted as a 'worker-discipline device'. Archival data now available allows for a basic account of the dynamics of the Gulag to be estimated. When this is combined with a dictatorship wishing to maximise the 'investible surplus' subject to an efficiency wage incentive constraint, what does it imply? That to secure resources for investment or war, consumption must be compressed; and making the Gulag harsher helps reduce incentive problems in the workplace. This is the cruel logic of coercion. But this economic rationale for the Gulag does not, we find, encompass randomised mass terror. Why did Stalin's system of coercion ultimately fail? The paper concludes with comparisons of Western and Soviet systems from an efficiency wage perspective.

JEL Classification: D82, P23, P26 and P27

Keywords: asymmetric information, efficiency wage, labour-discipline and Soviet Gulag

Marcus Miller
Department of Economics
University of Warwick
Coventry
CV4 7AL
Email: marcus.miller@warwick.ac.uk

Jennifer C Smith
Department of Economics
University of Warwick
Coventry
CV4 7AL
Email: jennifer.smith@warwick.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100440

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=117827

* We are indebted to colleagues Mark Harrison and Andrei Markevich for expert advice on Soviet economic history and data; to Chris Doyle, Gregg Huff and Robert Skidelsky (and seminar participants at Sheffield University) for their comments and suggestions; and to Tanya Kozin for research assistance. Marcus Miller acknowledges the financial support of an ESRC Professorial Fellowship, Grant No RES -051-27-0125.

Submitted 11 December 2007

Non-technical Summary

In the 1930s, when Western economies were laid low by mass unemployment, 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. Massive capital investment ensured no shortage of aggregate demand: the problem was how to compress consumption.

Supply-side incentives

Stalin was planning for great increases in productivity through rapid industrialisation and collectivisation. But how was he to motivate workers with low levels of skill – including 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? By extending coercion to promoting supply as well as creating demand is the answer explored in this paper, using an efficiency wage approach. For, as Stalin told the Party Congress in 1927, “Our plans are not prognoses, guess-plans, but *instructions*, which are *compulsory*”.¹

The growth of the Gulag

A brief overview and discussion of data now available on the custodial population in the USSR from 1917 to 1953, is followed by estimating a simple model of inflows and outflows into custody, including step-shifts to account for the Great Terror and for releases to the front during the Second World War. The idea is to combine this with an efficiency wage constraint and a dictator who wishes to maximise investible surplus.

Asymmetric information, property rights and ‘efficiency wages’

The efficiency wage theory we appeal to focuses on the problem of workers ‘shirking’ when information is asymmetric; such incentive problems were rife in a Soviet labour market featuring full employment. Appeal can be made to the ‘Coase theorem’ to show that either rewards or punishments can elicit effort, so long as property rights are appropriately determined. With wages held down to promote massive investment, the system had to depend on the stick rather than the carrot. But without high efficiency wages, how can incentives be preserved? From a Coasian perspective, shirkers could, in principle, be fined for failure to supply effort (and some such financial penalties were used); in practice, of course, workers on low wages simply cannot pay.

¹ Overy, 2004, p.406.

The logic of coercion

The alternative is to extend the command economy yet further – by using prison as an economic discipline device for shirking. Prisoners must be made to work, both to make them better citizens (‘corrective work’) and to augment the supply of goods; and if investment plans call for lower wages, prison conditions can be made more harsh. According to Gregory and Harrison (2005): “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”. This is the logic of coercion embodied in our efficiency wage analysis.

The framework we use is based Shapiro and Stiglitz (1984) for Western economies where workers are paid to put in effort, and failure to do so (shirking) leads to loss of employment and income. In our analysis, however, custodial sentences with corrective work in the Gulag replace spells of unemployment-on-benefit as the ‘worker-discipline device’.

The objective of maximising investible funds yields an equilibrium level of civilian employment, where the dictator exercises monopsony power to employ labour at the efficiency wage and no higher: those not so employed go to the Gulag. As an exercise in positive economics, the approach we take is brutally simplistic; but it focuses clearly on the issue of incentives, and provides equilibria where coercion replaces cash as an incentive device. The logic of coercion implies that investment can be further increased by reducing living standards in the Gulag. The predictions appear, moreover, broadly consistent with Sokolov’s (2003) argument that the solutions Stalin adopted involved a time-varying mix of coercion, moral suasion, and material incentives.

Russian Roulette?

Coercion as a discipline device helped Russia to industrialise at high speed – and to produce the arms needed to defeat Hitler: but, like the French Revolution, the regime of punishment became a monster. The archival data we draw on bear witness to the extent and longevity of the Gulag system developed and sustained under Stalin’s dictatorship – to say nothing of the settlements to contain more than two million kulaks.

The frenzy of punishment in the late 1930s may have served political purposes, but the economic rationale for the Gulag does not, we find, encompass randomised terror for political ends. Our efficiency wage model implies – as seems to be the case – that random terror is economically counter-productive: so too the random application of coercive labour laws. Stalin’s successors realised this, and the Gulag system and the harsh labour laws were dismantled soon after he died.

Why did the system finally fail?

How the command system survived for another quarter of a century is not something considered here. But the efficiency wage framework may be used to speculate briefly on why it finally failed. The inefficient use of the investible resources squeezed from the economic system, together with sluggish technical progress, meant that labour productivity in the Soviet system failed to catch up with the West. Furthermore, in an increasingly globalised environment where respect for human rights has been endorsed by all members of the United Nations including the USSR, the logic of coercion – that increased state spending requires wage restraint and productive efficiency calls for greater harshness of punishment – was much less acceptable. The effect of this on efficiency wages - together with a rise in monitoring costs as the structure of production shifted to services - put the command system under pressure. Add robust competition from the West, in military as well as in economic terms, and the logic of coercion could easily reach breaking point.

..... riches, poverty,
And use of service, none; contract, succession,
Bourn, bound of land, tilth, vineyard, none;
No use of metal, corn, or wine, or oil;
No occupation; all men idle, all;
And women too

Shakespeare *The Tempest* Act 2, scene 1

Introduction

At a time when Western economies were plagued by mass unemployment, Josef Stalin could rightly claim to have found a cure: a command economy with ambitious five year plans to catch up with the West by rapid industrialisation. Massive capital investment ensured no shortage of aggregate demand: the problem was how to compress consumption.

But those who would create Utopia without private property rights must confront the issue of incentives. This is evident from Gonzalo's vision of Utopia, cited above. For old Gonzalo the anticipated solution was natural abundance, produced "without sweat or endeavour".² But Stalin, for his part, was planning for great increases in productivity through rapid industrialisation and collectivisation. How was he to motivate workers with low levels of skill, including "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, p.315)?

How was Stalin to elicit the necessary 'sweat and endeavour' from his compatriots?

'Efficiency wage' theories may provide answers. Akerlof and Yellen (1990), for example, emphasise how worker motivation depends on whether employers are seen as good, and wages perceived as fair. This is the approach adopted to study incentives under Stalin by Gregory (2003), who uses it to explain the trade-offs involved in choosing between consumption and investment in the command economy. In Gregory's model, workers' effort depends positively on the wage (or consumption level) they receive, up to the point where they are paid the 'fair wage' and supply their 'full' labour effort. A dictator, wishing to maximise investment in the

² His companions were not convinced; nor, one assumes, was Shakespeare – shareholder of his theatre company and owner of the second most expensive residence in Stratford.

face of output constraints that force him to choose between investment and consumption, 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".³

The efficiency wage theory of Shapiro and Stiglitz (1989), on the other hand, focuses on asymmetric information and 'shirking'. Assuming the supply of effort is all or nothing, the worker is paid to put in effort, but failure to do so ('shirking') leads to loss of employment and income. Wages will need to exceed unemployment benefits by enough to preserve incentives for effort; but with imperfect monitoring of effort, incentive problems require the payment of 'efficiency wages' much exceeding the cost of effort-plus-benefit; and the maintenance of persistent unemployment as a 'worker-discipline device'.⁴ (Ironically, however, if unemployment acts successfully as a discipline device, there will be no shirkers among the unemployed, just those moving between jobs.)

Even a dictator has to solve endemic problems of asymmetric information,⁵ as Stalin was soon to learn. Although the first Five Year Plan was launched "with a wave of attacks on managers and specialists suspected of harbouring alien class sympathies", this was found to be "incompatible with the discipline drive, given their direct involvement in monitoring labour performance and implementing measures to designed to raise productivity"; and there was a sharp change of policy in 1931 (Acton and Stableford, 2005, p.316).

The Soviet system depended not so much on the carrot as on the stick (Harrison, 2002); but one can appeal to the 'Coase theorem' (1960) to show that either rewards or punishments can elicit effort, so long as property rights are appropriately determined. If labour power is effectively owned by the state, workers need not be rewarded for supplying effort. But without high efficiency wages, how are incentives to be preserved? From a Coasian perspective,

³ Gregory (2003) also mentions the possible use of forced labour to incentivise workers, the principle idea developed in this paper.

⁴ The loss of wages in suffering a spell of unemployment when caught and fired must be great enough to stop shirking; they show that the efficiency wage has to increase sharply as unemployment shrinks; and is also increasing in the level of non-incentive-related job losses.

⁵ The incentive problems arising from asymmetric information are central to Stiglitz's critique of the Soviet system in *Whither Socialism?* (1994).

shirkers could, in principle, be fined for failure to supply effort (and some such financial penalties were used); but in practice, of course, workers on low wages simply cannot pay.

Another solution is to extend the command economy yet further. This is the avenue we explore in this paper. It is an avenue that ultimately leads to the Gulag Archipelago,⁶ for the discipline device we consider is non-pecuniary deprivation – imprisonment in particular. Custodial sentences (with compulsory work) replace spells of low income and unemployment as an economic discipline device for shirking. As Gregory and Harrison (2005, p.740) note in their survey of allocation under dictatorship: “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”.

A brief overview and discussion of data now available on the custodial population in the USSR from 1917 to 1953, is followed by a simplified econometric account of inflows and outflows into custody. This account is then combined with an efficiency wage model developed in Sections 2 and 3, where the Shapiro and Stiglitz model of incentives is adapted to fit Soviet forms of coercion. The objective of maximising investible funds yields an equilibrium where the dictator exercises monopsony power to employ labour at the efficiency wage and no higher: but the logic of coercion implies that investment can be further increased by reducing living standards in the Gulag.

Broadly speaking, we examine the problem of maintaining incentives given that property rights reside with the state. Some observers, Kornai in particular, have argued that - in a wider analysis of the political economy of communism - the system of property rights was endogenous: that it was chosen so as to promote the power of centralised dictatorship.⁷ Tackling this wider issue is not the purpose of this paper;⁸ but we can say that the economic rationale for prison as a discipline device does not extend to Stalin’s unleashing of the Great Terror. How could the resulting mass executions and wholesale imprisonment help the economy? How could a dictator whose interests were to promote economic efficiency endorse a ‘monitoring rule’ where a success rate of only 5% was enough to justify denunciation?

⁶ Solzhenitsyn (1974).

⁷ Kornai (1992) argues that the collectivisation of agriculture, in Soviet Russia and elsewhere, showed political imperatives being given precedence over economic objectives.

⁸ Skidelsky (1995), Wintrobe (1998) and Gregory *et al* (2006) all provide interesting political economy perspectives.

As to why Stalin's system of coercion ultimately failed, we conclude with speculation based on the efficiency wage approach.

1. Data on 'corrective work' and on custodial population

While Shapiro and Stiglitz consider unemployment as a worker discipline device, it is clear that it could not perform that role in the Soviet system: by the early 1930s the Soviet government could rightly claim that unemployment was "liquidated" (Rogachevskaya, 1973).⁹ The proposition to be considered is that coercion not idleness was the discipline device in the Soviet case. But, as Sherlock Holmes warned Dr. Watson: "It is a capital mistake to theorize before one has data".¹⁰

Non-custodial punishment

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 48 per cent of all court sentences in 1935, for example (Getty *et al*, 1993, p.1020).¹¹ 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, p.1020; Sokolov, 2003, p.32). The several laws on labour discipline passed in the late 1930s and early 1940s increased the numbers given non-custodial sentences, but in relative terms the proportion sent to prison rose.¹²

Emergence of the Gulag Archipelago

⁹ There were some people in the labour force without work, but "under conditions of socialism, the condition of being without work (*nezaniatost'*) is not a synonym for 'unemployment'. It means only an interruption of work caused by reasons of a private character (family circumstances, changes of location)" Kotliar (1983, p.9).

¹⁰ In *A Scandal in Bohemia* (Conan Doyle, 1992, p.14).

¹¹ Solomon (1980) describes how the non-custodial sanction of what was then known as 'compulsory work' had been used extensively since the start of the Bolshevik regime. In 1923, only 20 per cent of those convicted in a criminal court were sent to prison; 25 per cent were sentenced to compulsory work (p.198). However, by 1926 the proportion sent to prison had risen to 40%, due to a change in the type of crimes coming before the courts and because judges began avoiding sentences of compulsory work that were not being properly carried out, in part because of the then high unemployment (p.204).

¹² From 20 per cent in 1930, to 37.8 per cent in 1934, to 55 per cent in 1938, and to more than two thirds in 1940 (Solomon, 1980, p.216), despite there being 1.7 million non-custodial sentences in 1940 (Getty *et al*, 1993, p.1020).

After Stalin and his allies took control of the Politburo in 1928-9, and after the decision to forcibly collectivise the peasants in 1929, numbers in custody began to rise inexorably. Chart 1 provides an overview of the numbers in custody over the years 1917 to 1953 (excluding settlements), with detailed figures and sources provided in Appendix 1. Note that in the text we use the term ‘prison’ to encompass the whole of the Gulag system, generally understood to include prisons, colonies and camps.¹³ (The data do not include figures for the mass executions characteristic of Stalin’s system – to be treated as symptoms of political repression and paranoia, not acts of labour-discipline, Wintrobe, 1998, Chapter 2).

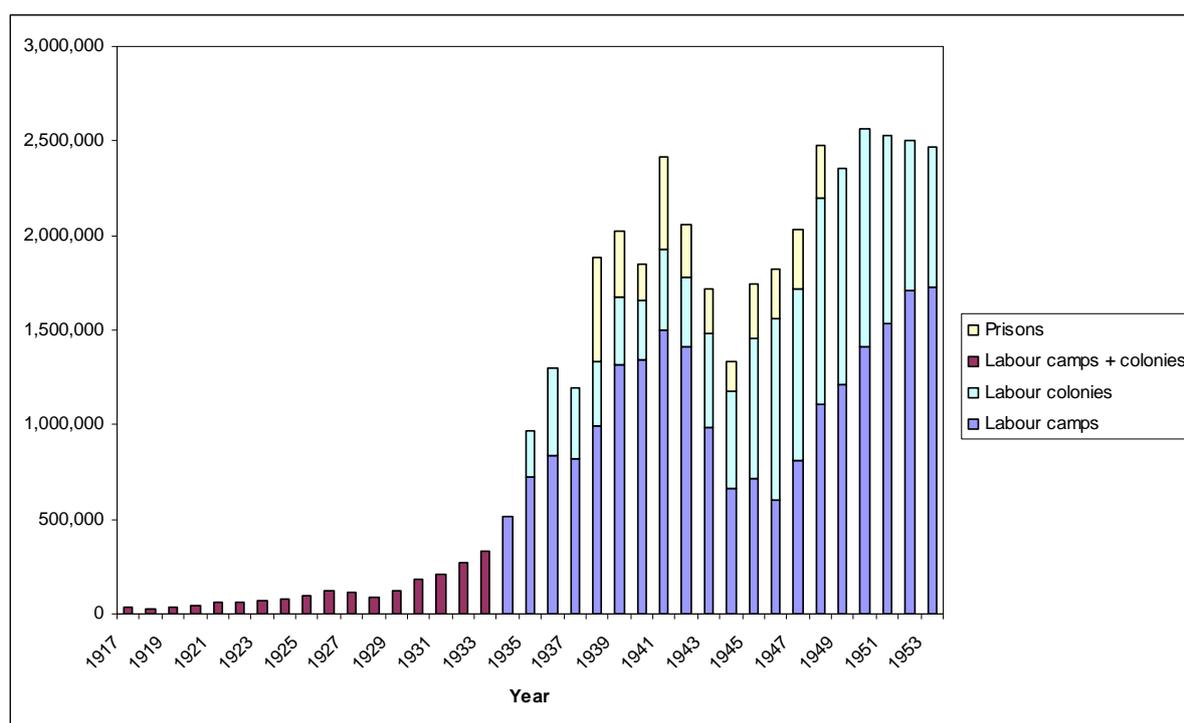


Chart 1: USSR custodial population, 1917-1953

Note: Data on prisons cover 1938-1948.

“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, p.435). The Law of Corrective Labour Camps of 1930 placed all camps and colonies in the control of the Gulag,

¹³ Gulag is an acronym for the Administration of Corrective Labour Camps and Colonies.

and the harsher sentencing after 1930 brought small-time crooks into the Gulag system so that by 1934, when the NKVD¹⁴ took charge of the camp system, around half a million were in custody (Overy, 2004). The NKVD tightened security and supervision, the possibility of escape diminished and the numbers imprisoned more than doubled in a couple of years. Thus the proportion of the working population imprisoned rose from 0.9% to 1.2% between 1934 and 1936 (when employment was 57.7m and 62.3m respectively).

The Great Terror

Imprisonment may act as a worker discipline device; but it was also used as an instrument of political power, with people being punished not for lack of effort but for ideological reasons. According to Lazarev (2003, p.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”.¹⁵

The ‘mass operations’ of the Great Terror lasted from July 1937 until November 1938.¹⁶ How the episode got its name – and the political drive behind it - becomes clear from the statistics. Not only did the number of arrests rise during the Terror, but the conviction rate also rose – from around one third in 1930 to 85 per cent in 1937 (Gregory *et al*, 2006, p.19). The result, as Chart 2 demonstrates for camps, was a huge rise in admissions to the prison system. There was no countervailing rise in releases – indeed, releases fell during the Terror – resulting in a 21 per cent increase in the Gulag camp population between January 1, 1937 and January 1, 1938, and an increase of 32 per cent the following year. Estimates vary, but even (conservative) data from the Soviet Archive show that, from a working population of 66 million,¹⁷ 1.4 million (over 2%) had been convicted by 1 November 1938, of whom about half were executed (Khlevnyuk, forthcoming).

¹⁴ People’s Commissariat of Internal Affairs – the secret police.

¹⁵ Furthermore, as Overy (2004) observes, to merit punishment under Stalin’s rule, it was not necessary to have committed an offence; it was enough that those in power thought you might do so on some future occasion.

¹⁶ During 1935-1936, Stalin had targeted the political elite, the three Moscow Show Trials enabling him to get rid of political rivals. In various communications and decrees of July 1937, 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 (Gregory *et al*, 2006).

¹⁷ In 1937.

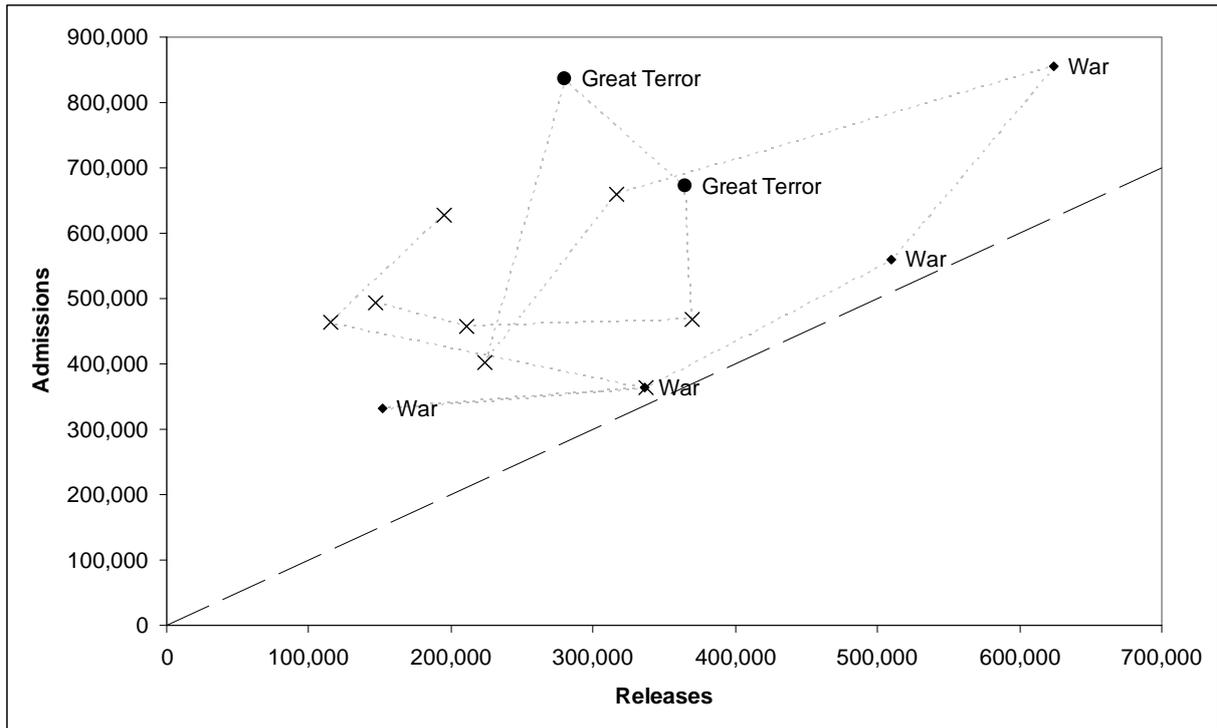


Chart 2 Admissions and releases: 1934-1947

Notes: Though strictly speaking it lasted from July 1937 until November 1938, the Great Terror is shown as 1937 and 1938. Likewise, War is shown as 1941-1945, though the War in Europe ended in early May 1945, and the War in the Pacific ended in August that year. That most observations lie above the 45 degree line of balance tallies with the inexorable expansion of the Gulag; but note that the dynamics of the prison population must also take account of deaths, executions and escapes not included in the chart.

Between 1937 and 1940, there was a four-fold increase in the number of political prisoners as Stalin purged civil society of counter-revolutionary elements (see Table 3 in the Appendix). At the height of the Great Terror, political prisoners accounted for around one third of the camp population. Political prisoners were charged under Article 58 of the Criminal Code, which allowed significant discretion in who to include among ‘enemies’.¹⁸

Stalin also used the administrative and legal system to increase labour discipline. To cope with absenteeism, lateness, drunkenness and high job turnover, tougher administrative measures were introduced in 1938,¹⁹ and, between 1939 and 1940, new laws turned absence from work,

¹⁸ Article 58 of the Criminal Code of the Russian Soviet Federal Socialist Republic was set out in 1927 to cover the arrest of suspected counter-revolutionaries (‘traitors’, ‘enemies of workers’, and ‘saboteurs’) and the categories were extended in 1934 and 1937. Flexibility arose in large part due to the offence of non-reporting, e.g. of anti-Soviet activities.

¹⁹ “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

tardiness, drunkenness and hooliganism into crimes²⁰ punished by up to four months in jail (Solomon, 1980, p.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, p.299); and there were over 4.5 million convictions during 1940-1942 (Sokolov, 2003, Table 2.1, p.28).²¹

War and Post War

From a peak of over two million during the Great Terror, the numbers in custody fell to a million and a third by 1944. This was in large measure due to a step increase in the release rates connected with the war effort, see Chart 2. (Almost one 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, p.51), due to scarce rations and the fact that the most able-bodied had been sent to war.

But after the war was over, the custodial population rapidly resumed its upward march, reaching a plateau of two and a half million in 1948. Numbers stayed at this level until Stalin’s death in 1953 when more than half of all detainees were released. Pre-war labour discipline laws were retained after the war, but labour turnover remained a problem, reaching 34 per cent in light industry and 64 per cent in construction (Sokolov, 2003, p.37);²² and a further decree combating mobility was issued in 1948. 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

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, p.25).

²⁰ “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. The October 1940 reforms of vocational education raised the term of obligatory work after graduation to four years and prohibited voluntary departures” (Sokolov, 2003, p.25).

²¹ Not all of these served prison sentences: according to Sokolov (2003, p.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 5.3 million trials for such offences (Sokolov, 2003, pp.27-28), which represents 2.8 per cent and 4.3 per cent of the workforce, respectively.

²² Living conditions were poor and were exacerbated by a famine in 1946-47. “A female worker in a Moscow plant wrote: ‘We worked hard throughout the war; we awaited the victory and counted on better conditions. The opposite occurred. They lowered our salaries and we receive pennies. It is time to think about the workers’.” (Sokolov, 2003, p.34).

discipline law.²³ Labour restrictions were eventually reduced in 1951,²⁴ but they were only fully abolished in 1956, after Stalin's death.

The growth of the Gulag Archipelago: an econometric sketch

The dynamics of the prison population depends on inflows (reflecting the rate of imprisonment) versus outflows (reflecting releases, escapes and deaths in custody). We use data on these flows during Stalin's era to see what they reveal about the dynamics of the prison population – and its steady state equilibrium.

On the assumption that the annual inflow rate π and outflow rate ρ^+ (to denote releases *plus* escapes and deaths) are constant proportions of prison population P and labour force L , respectively, the annual change in the prison population over year t , ΔP_t , will be given by $\Delta P_t = \pi L_{t-1} - \rho^+ P_{t-1} + \varepsilon_t$ where we assume ε_t is i.i.d. Normal and $\rho^+ > 0$.²⁵

Fitting this to the data over the Stalin era, 1928-1953, but including zero-one step dummies to account for the Great Terror (1937-1938) and the Great War (1941-1945), gives the following parameter estimates:²⁶

$$\Delta P_t = 0.006L_{t-1} - 0.209P_{t-1} + 368,384\text{Error}_{t-1} - 148,545\text{War}_{t-1} \quad \bar{R}^2 = 0.20; N = 25$$

(0.002) (0.100) (240,863) (165,915)

These estimates imply that a little over half of one percent of the civilian labour force were incarcerated each year, and around one fifth of prisoners were released (or died in custody), these flows averaging just under 400,000 annually. Though the relevant shift dummies are statistically not well-determined, they reflect the marked rise in prison admissions during the Terror, and of releases during the War.

Dividing through by workforce N , and assuming N to be constant, gives

²³ The relative severity of punishment for these offences rose, as the numbers fined fell by half (Sokolov, 2003, pp.38 and 41).

²⁴ By a decree of the Praesidium of the Supreme Soviet of July 14 '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, p.38).

²⁵ 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.

²⁶ Standard errors are given in parentheses. The model was also estimated using interactions between the Terror and War dummies and each of the other variables. Results accorded with priors: estimated inflow and outflow rates were significantly higher in the two periods (and much higher during the Terror), and the steady-state annual flow at other times was approximately 180,000. But due to the small sample size, we prefer to report results based on simple shift dummies.

$$\Delta p = \pi - (\pi + \rho)p$$

where $p=P/N$ and $\Delta p = \Delta P/N$. So in steady state, where $\Delta p = 0$,

$$p = \pi / (\pi + \rho^+).$$

The estimated coefficients imply that in equilibrium $p = 0.006 / (0.006 + 0.209) = 0.029$; in other words roughly three per cent of the working population would be in prison in steady state. During Stalin's rule the working population (approximated by employment) averaged 67 million, implying a prison population of 1.95 million – an equilibrium which broadly matches the figures after the NKVD took charge in the mid 1930s (see Appendix Table A1).

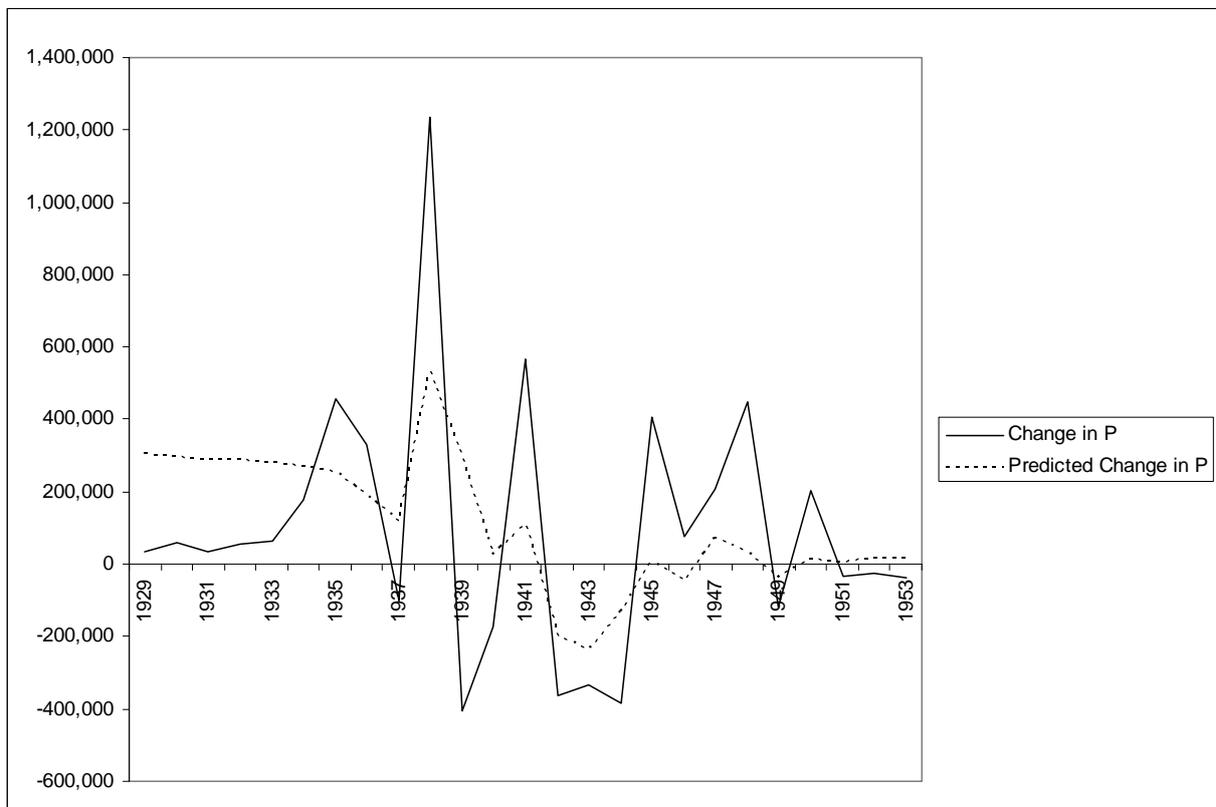


Chart 3 Changes in custodial population: actual and predicted

Chart 3 shows that predictions from the estimated equation broadly track the major swings in net admissions, though prison inflows prior to the NKVD taking in 1934 are substantially over-predicted. (Note, however, that this was when Stalin forcibly re-settled the kulaks;²⁷ and the custodial figures we use exclude such resettlements.)

²⁷ Approximately 2.3 million kulaks were sent for resettlement between 1929 and 1940; of whom 1.8 million were sent into exile by the end of 1931 (Ivnitskii, 2004, p.43; Zemskov, 1990).

2. Prison as an incentive device

A simple characterisation of the dynamics of the Gulag system has been provided: but its economic rationale will depend largely on its effect on the incentives of those in civilian employment outside prison. How to analyse such incentives?

We begin with Shapiro and Stiglitz's account of how shirking is checked under capitalism. It has three salient characteristics: that the punishment for being caught shirking is to lose one's job; that efficiency wages paid to the employed are a lot higher than effort-plus-benefits (as paid to the unemployed – who are not required to work); that this premium rises sharply as employment levels increase and unemployment falls. (The reason for the rising premium is that the punishment involved in losing one's job is diminished by the short unemployment duration rates prevailing at low levels of unemployment.) Analytical detail of the No Shirking Constraint (NSC) that they obtain in this framework is presented in Appendix 2.

As Wintrobe (2000, p.28) remarks, however, 'The prospect of unemployment is not the only possible discipline device'; and, in the context of a two-sector model of labour markets,²⁸ he notes that 'Dismissal poses a threat as long as the worker can fall into the bad jobs sector. Such a prospect is sufficient to discipline the labour force.' Here we consider a Soviet alternative where there is no unemployment; but the punishment for those caught shirking is not a 'bad job' but being sent to a labour camp, where work is compulsory.

Efficiency wages and 'Dire Punishment'

Efficiency wages where shirking is treated as a crime meriting imprisonment will surely depend on prison conditions and on duration of punishment. To capture the psychological impact of being 'sent to Siberia', we begin with the case where imprisonment is seen as the end of normal life (labelled 'dire punishment', a term used in repeated games to denote a state from which there is no transition). As Ertz (2007, p.27) puts it: "For individuals sentenced by the Stalinist political or criminal justice, ...their chances to turn into 'Soviet people' were, if not zero, then at least much lower than for the rest of society".²⁹

²⁸ That of Bulow and Summers (1986).

²⁹ Even after release, an ex-con would not be able to participate normally in society: for example, on release, political prisoners had to sign a paper stating that they would never again engage in counter-revolutionary activity, were forbidden to live in major cities and had to report to a police station of the NKVD for years afterwards (Overy, 2004, p.634).

Let w denote the real wage and e denote effort while working, so the welfare of one who works is simply $w - e$, i.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 hard labour. The benefit of working *versus* being in prison will therefore be $w - e - \gamma$. (Note that γ will be negative if being in prison is worse than being paid just enough to put in effort voluntarily outside prison.)

In the dire punishment case, where incarceration is treated as permanent and r denotes the discount rate, the (continuous time) ‘no shirking condition’, NSC, is

$$e \leq q(w - e - \gamma) / r \quad (1)$$

i.e. the benefit of saving on effort for one period must match the risk of losing one’s job and being imprisoned for ever.

The ‘efficiency wage’ is where the two are exactly equal, i.e.

$$w^* = e + \gamma + re / q, \quad (2)$$

so it rises with the level of effort and the interest rate; and falls with the efficacy of monitoring and with the harshness of prison conditions. Crude as it may be, this can offer insights into aspects of the Soviet system.

(a) Shooting the monitor

During the first Five Year Plan, managers and specialists were harassed and imprisoned for ideological reasons, reducing the efficacy of monitoring. Equation (2) shows that a fall in the probability of detection q has the immediate effect of raising the efficiency wage, with potentially serious incentive effects (considered in more detail below).

It seems that by 1931-1932 Stalin had learnt this lesson, for the policy was then changed. “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, p.316).

(b) Stakhanov's shovel

After the celebrated example of Stakhanov, who in 1935 mined far more coal per hour than the norm,³⁰ many managers hoped that others would follow his lead and produce more coal for the same wage. But the formula for the efficiency wage confirms that “harder work deserves a bigger share of the pie and that the higher wages of Stakhanovites should be passed on to ordinary workers” who were also working harder (Gregory, 2003, p.106).³¹ In fact “ordinary workers interpreted the Stakhanovite movement as a plot to extract more work for the same wage”, and there was “no perceptible advance in labour productivity during or after the Stakhanovite movement, despite the fact that some Stakhanovites raised labour productivity substantially” (Gregory, 2003, p.105).

(c) Promises, promises

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. Can the efficiency wage be restrained by “visions of a brighter future” (Gregory, 2003, p.97)?

That workers might indeed be willing to accept a lower efficiency wage in return for future increases can be seen by augmenting the right hand side of (1) by the term $J \exp^{-rT} / r$, where J is capitalised value of the promised ‘jump’ in wages at time T , appropriately discounted. The efficiency condition becomes:

$$e = q(w - e - \gamma + J \exp^{-rT}) / r ; \quad (1')$$

and the efficiency wage consequently falls by the present discounted value of the jump, so:

$$w^* = e + \gamma - J \exp^{-rT} + er / q . \quad (2'')$$

Promises of a brighter future can, in principle, maintain incentives despite a cut in current wages. But for incentives to be preserved, promises must be credible: and credibility became strained as the Plans failed to deliver.

³⁰ He cut 102 tons of coal in 5 hours 45 minutes, beating the ‘norm’ of 7 tons by a factor of over 14 (Gregory, 2003, p.103).

³¹ In fact, because of imperfect monitoring, the efficiency wage rises more than one for one with effort, as can be seen by differentiating (2) with respect to effort to obtain: $\partial w^* / \partial e = 1 + r / q$.

3. Efficiency wages in coercive equilibrium

During the Great Terror and afterwards, “Such a wide range of behaviors was criminalized that virtually every worker became liable to prosecution for something” (Gregory and Harrison, 2005, p.739). The random nature of punishment is exemplified by the ‘five per cent rule’ for denunciation, which Stalin expressed as follows: ‘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, p.18). This seems likely to increase terror more than promote detection,³² 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*, 2006, p.18).

For workers faced with increased uncertainty concerning their liberty, however, such random denunciation (and consequent punishment) will have the opposite effect of brave promises; it will reduce the value of a job and increase the efficiency wage. For, assuming that the average worker treats these risks of imprisonment as a hazard that arrives randomly at rate πdt independent of effort supplied, the flow of net earnings for a job must be discounted at a higher rate, to allow for the risk of random ‘state transition’ as well as the passage of time: consequently the rate of discount in equation (1) increases from r to $r + \pi$.³³ As the value of a job falls, so the efficiency wage rises to become

$$w^* = e + \gamma + (r + \pi)e/q. \quad (3)$$

Thus, in terms of the efficiency wage, random threats of dire punishment militate against promises of a bright future.

In equilibrium where wages satisfy this condition, however, no-one will be shirking, so admissions to prison will be for political and other reasons. As indicated by the econometrics

³² 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*, 2006, p.22, citing Vatlin, 2004).

³³ 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 below and by random applications of labour law. As Gregory and Harrison (2005, p.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”.

in the previous section, the dynamics of prison population and its steady state will depend on the prospect of release (and of death in prison, which occurred on average at about one sixth of release rate).³⁴ The effect of incorporating the prospect of release from prison back to a normal working life at the rate ρdt ³⁵ – so that prison is no longer an absorbing state – is to further increase the rate of discount on the RHS of equation (1) and to increase the efficiency wage. A higher probability of death in prison will have the opposite effect by reducing the expected value of life in the Gulag. So the efficiency wage may be written

$$w^* = e + \gamma + (r + \rho + \pi)e/q \quad ^{36} \quad (4)$$

where γ has been reduced accordingly.

By treating the rate of admissions as endogenous, as required to maintain the prison population constant, we finally obtain the incentive effect relating the efficiency wage to the fraction of the labour force in gaol.

Proposition 1 There is a multiplicity of coercive equilibria in which the efficiency wage moves inversely to the prison population.

The specific relationship, derived in Appendix 3, involves substituting for ρ in (4) to give:

$$w^*(p) = e + \gamma + (r - \delta + \pi/p)e/q \quad (5)$$

The efficiency wage so defined is shown as the schedule labelled NSC in Figure 1(a), sloping upwards from its intercept with the vertical axis (where p is unity so (5) yields the constant $w^*(1) = e + \gamma + (r - \delta + \pi)e/q$); and becoming vertical when the release rate reaches its maximum value $\bar{\rho}$, as indicated in the lower panel.³⁷

³⁴ Our simplified treatment does not include the impact of executions or escapes.

³⁵ Note that the lower the release rate, the longer the expected spell in prison, so one could draw an equivalence between a model with stochastic release and one with a determinate prison sentence, where ρ is inversely related to the length of sentence.

³⁶ The other method of getting out of prison alive – escape – will have an effect similar to that of release. The possibility of escape from the Gulag diminished after 1934, when the Soviet secret police (NKVD) took over the whole of the camp and colony system. Nevertheless, archive data indicate that between 1934 and 1953, 378,375 escapes were attempted; only 38 per cent of these succeeded, however (Getty *et al*, 1993, p.1041).

³⁷ Evidence shows varying release rates: over the period 1934-1952, they varied between 15 and 45 per cent of Gulag inmates. It seems realistic to bound the release rate in this way: but it is not essential to our argument.

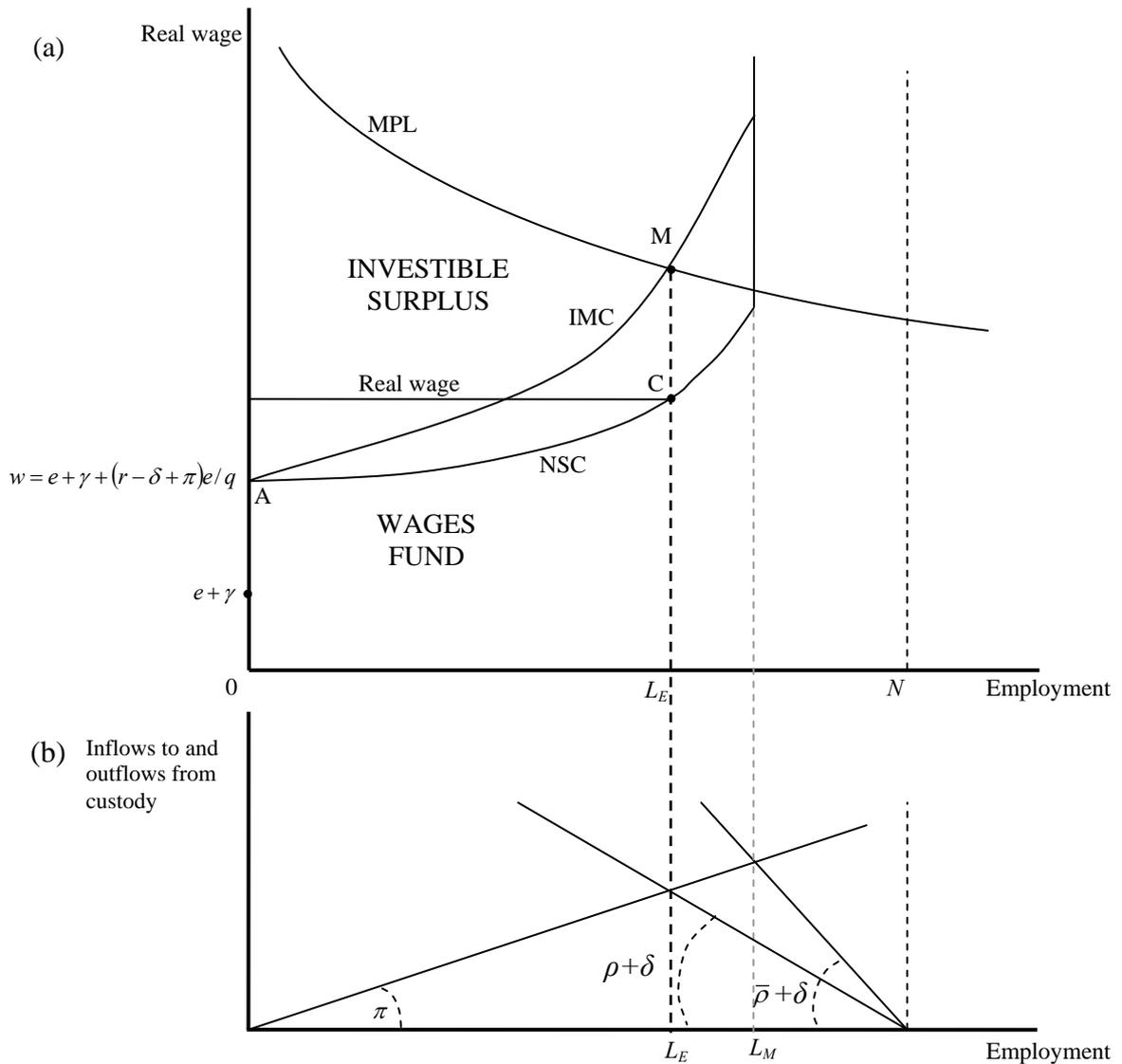


Figure 1: Investment maximising equilibrium

To complete the description of macroeconomic equilibrium one has to ask how wages will be determined. According to Skidelsky (1995, p.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 is simple enough:

Proposition 2 The ‘technical requirement’ for limiting the wages fund so as to maximise investment is that the state act as monopsonist in the market for civilian labour.

That is to say L and w are chosen so as to maximize the surplus value available for the state (to invest in future growth, for example) consistent with the willingness of labour to work. As is shown in the Appendix, this implies that wages are forced down to their efficiency level and that the demand for labour is restricted so that the efficiency wage lies below the marginal product of labour. Specifically it is shown that

$$F_L = w^* + Lw_L^* + s \quad (6)$$

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 s is the ‘surplus value’, if any, extracted from the prison population.³⁸ Thus the MPL must lie above 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.

[This optimising condition 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 Curve which determines the level of civilian employment. From the definition of the NSC in equation (5) above and the definition of p it follows that

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

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

$$F_L = w + Lw_L^* = e + \gamma + (r - \delta + \pi / p)e / q + e\pi(1-p) / (qp^2) = e + \gamma + (r - \delta + \pi / p^2)e / q \quad (7)$$

Note that the marginal cost of labour, shown as the steeper curve labelled IMC in Figure 1(a), takes the same form as the NSC (the average cost), except that the parameter π is divided by p^2 instead of p .]

The nature of coercive equilibrium may be summarised as in Figure 1 where the intersection of the Investment Maximising Curve with the Marginal Product of Labour at point M determines the level of employment and production; and the wage is pushed down to lie on the No Shirking Constraint at point C. Since the Wages Fund is determined as the product of the real wage C and civilian employment L_E , the remainder of output – the area above the real wage line – will be available for Investment.

³⁸ If the Gulag was designed to be simply self-sustaining and no more, then $s = 0$. This is the ‘neutral’ assumption made in the algebra to follow.

The steady-state nature of the equilibrium is indicated 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 and deaths *times* the custodial population, $(\rho + \delta)P$, where the custodial population is $P = N - L_E$, and the parameter ρ is treated as endogenous.

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. The state commands a goodly share of national resources, but wages are pushed down to efficiency levels. Ironically, the outcome for labour is as if it faced a greedy capitalist who wanted to maximise profits and had the monopsony power to do so. More than that, the employer can manipulate the living and working conditions for those not in civilian employment too.

But there were problems. Labour demand driven by massive investment made it difficult to administer lower wages as long as workers could change jobs. As Acton and Stableford (2005, p.317) note: “Workers retained a rudimentary amount of autonomy and bargaining power ... The effect was to compel managers to compete with each other to secure and retain labour ... Certainly individual workers were subject to arbitrary mistreatment, but if the management of one factory was too heavy-handed, a worker could move on and sign up in another where the demands made on him or her were less exacting, wages higher, or conditions better”. The result was that, for a time, “there was a thriving labour market that moved workers from enterprise to enterprise, even when coercion was most intense” (Gregory and Harrison, 2005, p.748). As is clear from Figure 1 where the MPL exceeds the real wage, managers determined to fulfil production targets can poach labour from elsewhere by paying more than the administered wage, with soft budget constraints giving them the freedom to do so.

In the tough labour laws of the late 1930s, Stalin circumscribed the freedom of labour to move from job to job. Thus was the pre-revolutionary prophecy of Karl Kautsky fulfilled. Writing about the class struggle in 1910, he had stated: “socialist production is irreconcilable with the full freedom of labour, that is with the freedom of the labourer to work when, where and how he wills”.³⁹

³⁹As cited in Nove (1991,p.210) and Skidelsky (1995, p.100).

5. Shifts of coercive equilibrium

Making the Gulag harsher

We have seen how investment is maximised conditional on a given standard of living in the Gulag. But what if the dictator can vary prison conditions as well? Will he not be tempted to do whatever eases the incentive constraint? Indeed we find that:

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

A derivation of this result is provided in the Appendix; and it can be seen clearly in Figure 2(a), 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 C to C'.

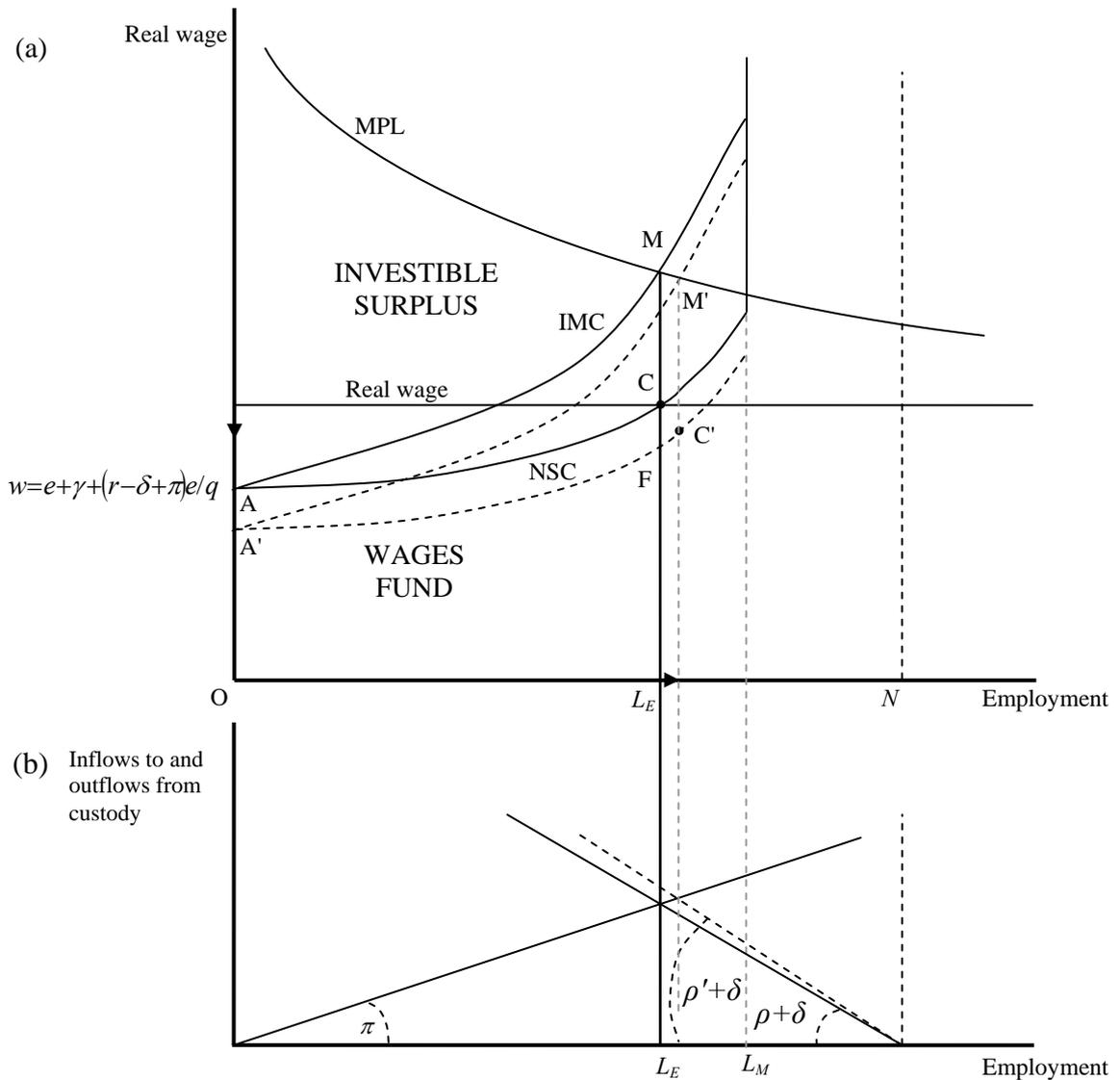


Figure 2: Harsher Gulag, more investment

The fall of the efficiency wage obviously increases investible funds at the pre-existing level of production. But the corresponding cut of the IMC curve 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 yet increase civilian 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 portrayed indelibly in *One Day in the Life of Ivan Denisovich*.

The framework outlined here is, we believe, consistent with Sokolov's (2003) view that the strategy adopted involved a time-varying mix of coercion, moral suasion, and material incentives. The availability of consumer goods to provide material incentives was limited by the emphasis placed on investment in heavy industry by Soviet economic planning. The first Five-Year Plan motivated workers during 1928-1933, with current low wages coupled with promises of a brighter future. The removal of 'class enemies' did not bring material rewards – indeed by reducing monitoring it imperilled production and called for a major change of policy. Stakhanovism during the mid-1930s encouraged ever greater individual productivity, but did not succeed in the aggregate.

Like unemployment in the West, the prison system did have a rationale in terms of lowering incentive constraints on production. But does it provide an economic justification for the Great Terror?

Increasing Terror

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 – typically induces paranoia. “The insecurity of the dictator results in excessive repression”, he observes, “Stalin is the classic modern example”. Janos Kornai (1992, p.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.

A chilling political rationale for the Great Terror of 1937-1938 – in terms of maintaining a supportive constituency by large-scale elimination of 'enemies' – has recently been provided by Gregory *et al* (2006). How effective this proved in terms of cementing Stalin's power is not something we investigate here. But what our model of incentives shows is that the creation of random terror on a massive scale cannot plausibly be attributed to economic objectives. This seems to accord with the facts. As Acton and Stableford (2005, p.386) observe: “If a major aim of the Great Terror was to overcome endemic waste and poor-quality production, to remedy the inherent malfunctions of the Stalinist form of 'planning', or to compel regional and local officials to obey Moscow to the letter, it ... failed dismally”.

The Great Terror surely reduced the civilian labour force: but what of the effect on incentives?
 The logic used to see the effect of prison conditions leads to the conclusion that:

Proposition 4 Random terror is economically counter-productive: increasing the rate of incarceration, the parameter π , will raise the prison population and the efficiency wage and reduce the investible surplus.

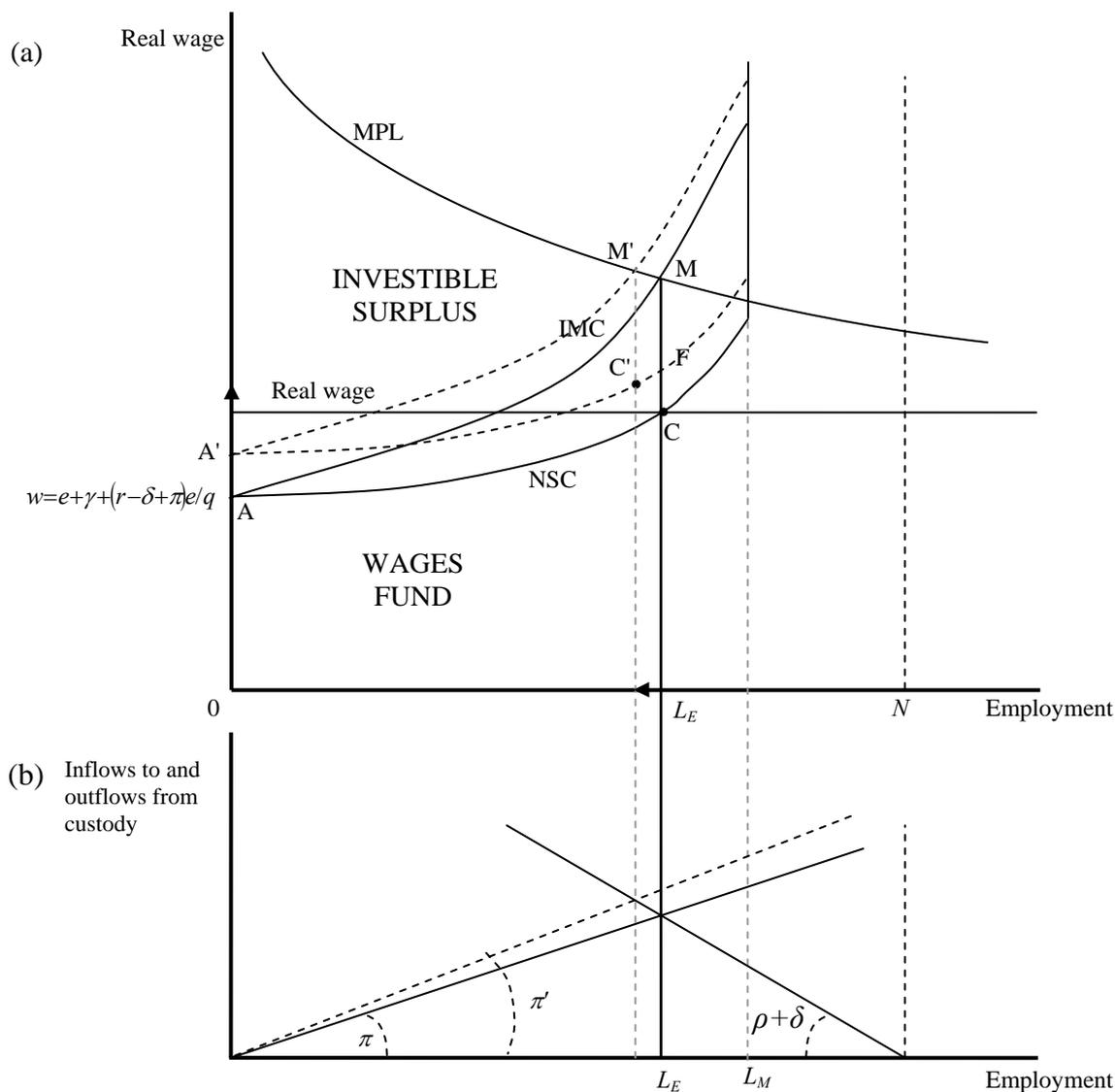


Figure 3: Greater terror, less investment

Formal details are in the Appendix; but the essence of the argument is shown in Figure 3 where unleashing terror is captured by an increase in the parameter π (the perceived threat of random incarceration). The effect is to shift the common intercept of the NSC and IMC from A to A' , contracting the equilibrium level of production from B to B' , raising wages from E to E' and

reducing the surplus available for investment and growth. This effect is the opposite of that shown in the previous figure, where greater iron discipline in the Gulag increased the real resources available for investment.

That Stalin was willing to sacrifice economic goals for increased political control seems clear from his enunciation of the infamous 5% rule. How could 95% false positives conceivably be justified in the name of economic efficiency? As the earlier citation from Wintrobe suggests, the unleashing of randomised terror and of incitements to widespread misreporting appear not as acts of labour discipline but as symptoms of paranoia and repression.

This is consistent with the fact that, even before Stalin's death, Lavrenty Beria (Stalin's secret police chief) and his MVD (Ministry of Internal Affairs) were planning large scale modifications to the Gulag system. Tikhonov (2003) outlines various MVD plans, drawn up from around 1949 onwards, to convert Gulag sentences into exile to remote regions to work on MVD industrial and construction projects (at lower wages than free workers). Exiled ex-prisoners would be allowed to bring their families and set up home. The aim of these proposals was not humanitarian, but was designed to meet production goals and reduce expenditure on prisons: at given levels of prison productivity, the population had become too large.⁴⁰

6. Why did Stalin's coercive system collapse?

Though it is not the object of this paper, it is tempting to look beyond Stalin's years of power to see if the framework we use offers any clues as to why the system he created finally collapsed. To this end, we briefly compare the two systems, of Western capitalism and Soviet coercion, in terms of incentive constraints for labour, criteria for wage determination, and of demand management, see Figure 4 (where the labour force is normalised at N).

⁴⁰ "The Gulag became a drain on the economy and the state budget as it filled with victims of the Great Terror and then with returning Soviet POWs. The 1953 amnesty derived primarily from the bureaucratic interests of the MVD itself. An external event – Stalin's death – merely provided an excuse for the radical reform, which had been desired by the MVD and Gulag administration itself for many years" (Tikhonov, 2003, p.73).

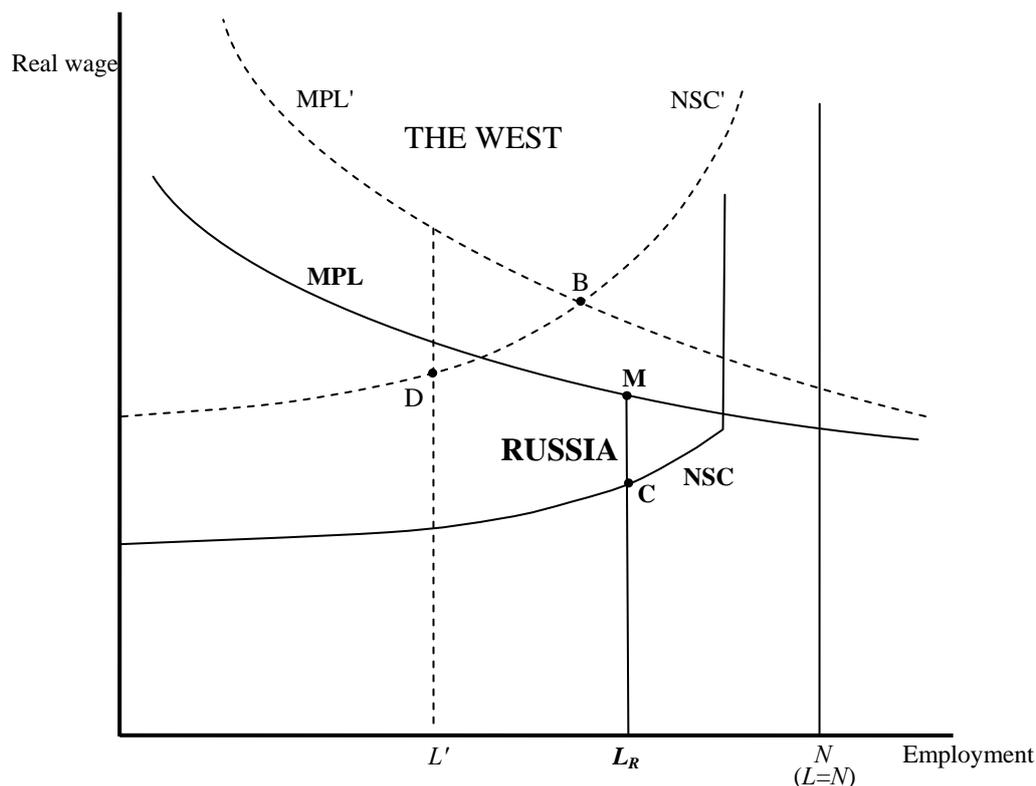


Figure 4 Efficiency wage equilibria in Russia and in the West

Shown in bold is the Russian case, with the No Shirking Constraint based on imprisonment. On the Marginal Product of Labour schedule, let the point M be where the marginal cost of labour⁴¹ matches its marginal productivity, defining the employment level L_R for a dictator wishing to maximise the resources available to the state for investment, military spending, or whatever the dictator decrees. Only the lowest incentive-compatible level of wages will be paid: so the coercive equilibrium lies on the No Shirking Constraint, below M at point C, generating extra investible resources of MC *times* the volume of employment.

In the Western case, with unemployment as a labour-discipline device, the incentive constraint is higher (as shown by the dotted line NSC'); and so too is labour productivity (MPL'). But we need to consider two possible equilibria, shown at D and B, depending on whether there is a demand constraint or not.

We start in the 1930s with the West mired in the Great Depression, diagnosed by Keynes as demand failure. With a demand constraint (shown by the dotted vertical line in the figure)

⁴¹ Shown as IMC in Figure 1, for example.

limiting the quantity of output that can be sold,⁴² let labour market equilibrium be at D, with high unemployment $N - L'$ and wages depressed to the lowest incentive-compatible level. In this quantity-constrained equilibrium profits are maximised subject to a demand constraint, as would be consistent with weak labour unions.

Point C symbolises Stalin's contemporaneous experiment in coercion. How does it compare with what Western capitalism had to offer? The productivity of labour is lower than in the West due to the low initial level of capital per head in Russia at the beginning of industrialisation; and the compression of real wages is a sign of the effort to catch up with the West by rapid capital accumulation. Real wages are not high in either system, but the massive investment in Russia carries promises of a brighter future. Compared with the West in Depression, higher employment in Russia is relatively attractive for those in jobs: but for those not in jobs, the Gulag replaces the unemployment queue.

Let time move forward some fifty years – to 1984 say, when Shapiro and Stiglitz publish their paper. Let the competitive equilibrium they describe represent the West after it has learnt the art of macroeconomic stabilisation: at B there is no demand failure; and output is maximised subject to the incentive constraint imposed by asymmetric information.⁴³ How does Stalin's system compare now? Assuming – as George Orwell had foretold – that Russia remains in equilibrium at C, it offers lower wages, lower output – and widespread coercion.

Figure 4 takes as given that labour productivity in the West continues to run ahead of that in Russia despite fifty years of squeezing consumption to release resources for investment and growth. Is this plausible? It would be if the resources so painfully extracted from the Russian people were not invested efficiently;⁴⁴ and if the bureaucratic and centralised system based on fear failed to match the incentives in the West for innovation and continuous 'technical progress', as argued publicly by the Russian physicist Sakharov and colleagues, for example.⁴⁵

⁴² As discussed at the end of Appendix 2.

⁴³ The use of the same MPL' curve is, of course, simply for convenience.

⁴⁴ "The system's chief deficiency was the unprofitability of its investments. Although based on the principle of investing more of its national income than a capitalist society could achieve, the regime could not make enough from its investments to support itself" (Skidelsky, 1995, p.107)

⁴⁵ "It is no accident that all the great scientific and technological discoveries of recent times – quantum mechanics, new elementary particles, uranium fission, antibiotics, and most of the new, highly effective drugs, transistors, electronic computers, the development of highly productive strains in agriculture, the discovery of other components of the 'Green Revolution', and the creation of new technologies in agriculture, industry, and construction – all of them happened outside our country" Sakharov (1975, p.31). See also the 1970 Manifesto, Sakharov (1974, Chap. 3).

But the coercive system faces yet further challenges. The spreading doctrine of Human Rights makes repression less acceptable, raising the efficiency wage in Russia⁴⁶; and Western spending on an ambitious and expensive arms race gives the Russian government added reason to compress wages - to pay for arms as well as investment⁴⁷. With upward pressure on the efficiency wage and a strict limit on resources available for consumption, will the time not come when the incentive constraint is violated and the command economy collapses (everybody 'shirks')?

6. Conclusions and qualifications

At a time when Western economies were prostrated by mass unemployment, Stalin's command economy achieved full employment. Though he may have been driven by the urge to challenge Western industrial supremacy, his methods owed little or nothing to Western ideas: Keynes had yet to write the *General Theory* (1936), for example.

As command had cured unemployment, so it was expected to solve the problem of incentives: that is the logic of the model developed here as an exercise in positive economics. The approach we take, like Shapiro and Stiglitz's capitalist equivalent, is brutally simplistic. It may lack the psychological subtlety of Akerlof and Yellen's doctrine of fair wages; but it focuses clearly on the issue of incentives, and generates equilibria where coercion replaces cash as an incentive device. It would nevertheless be of interest to extend the analysis to allow for variable effort;⁴⁸ and it should also be possible to combine both prison and unemployment as punishment devices.

Coercion as a discipline device helped Russia to industrialise at high speed – and to produce the arms needed to defeat Hitler: but, like the French Revolution, the regime of punishment became a monster – with mass executions of those perceived to be enemies of the state. The archival data we draw on bear witness to the extent and longevity of the Gulag system

⁴⁶ The shifting of production to the service sector, where monitoring is more difficult, will also have raised efficiency wages, as Harrison (2002) points out.

⁴⁷ In the words of Skidelsky (1995, p.111): "The command economy - the people - had to be flogged even harder. But neo-Stalinist incentives no longer worked."

⁴⁸ As Shapiro and Stiglitz (1984) suggest, "Including effort as a continuous variable would not change the qualitative results" (p.435).

developed and sustained under Stalin's dictatorship – and that is without taking into account settlements to contain more than two million kulaks.⁴⁹

Skidelsky (1995, p.103) argues that “the Stalinist command economy was not a technical solution to the economic problem of inadequate saving and investment: it was a device for maximising and perpetuating the power of the state”. Our analysis shows, however, that there is an economic rationale for a Gulag system as a labour-discipline device; and it shows that there were economic incentives to make it harsh. But the frenzy of punishment in the late 1930s (and subsequent randomness in the application of coercive labour laws) cannot be rationalized on economic grounds.

Stalin's successors realised this, and the Gulag system and the harsh labour laws were dismantled soon after he died. How the command system survived for another quarter of a century is not something considered here. But a command economy will surely ‘fall apart’ when incentives fail: “Some scholars have argued that coercion was the glue that held the Stalinist economy together: when coercion failed, or was abandoned, the economy fell apart” (Gregory and Harrison, 2005, p.735). Was it the relative success of market economies – together with Star Wars and the spread of Human Rights – that finally destroyed the coercive system Stalin created?

⁴⁹ Stalin's youth as a street fighter in Georgia (Sebag Montefiore, 2007) lends credence to the idea that he saw brute force as tool of social policy.

References

Acton, Edward and Tom Stableford (2005), *The Soviet Union: A Documentary History, Volume 1: 1917-1940*, Exeter: University of Exeter Press.

Akerlof, George A and Janet L Yellen (1990), "The fair wage-effort hypothesis and unemployment", *Quarterly Journal of Economics*, 105 (2), 255-283.

Bacon, Edwin (1992), "Glasnost' and the Gulag: new information on Soviet forced labour around World War II", *Soviet Studies*, 44 (6), 1069-1086.

Bacon, Edwin (1994), *The Gulag at War: Stalin's forced labour system in the light of the Archives*, London: Macmillan.

Bulow, Jeremy and Lawrence Summers (1986) "A theory of dual labor markets with applications to industrial policy, discrimination and Keynesian unemployment", *Journal of Labor Economics*, 4, 376-414.

Coase, Richard H. (1960), "The problem of social cost", *Journal of Law and Economics*, 3, October, 1-44. Reprinted in: R. H. Coase, *The Firm, the Market and the Law*, Chicago: University of Chicago Press, 1988.

Conquest, Robert (1994), "Communication to the Editor", *American Historical Review*, 99 (3), 1038-1040.

Doyle, Arthur Conan (1992), *The Adventures of Sherlock Holmes*, London: Reader's Digest Association.

Ellman, Michael (2002), "Soviet repression statistics: some comments", *Europe-Asia Studies*, 54 (7), 1151-1172.

Ertz, Simon (2005), "Trading effort for freedom: workday credits in the Stalinist camp system", *Comparative Economic Studies*, 47 (2), 476-491.

Ertz, Simon (2007), "Making sense of the Gulag: analyzing and interpreting the function of the Stalinist camp system", PERSA working paper, 10 August.

Getty, J Arch, Gabor T Rittersporn, and Victor N Zemskov (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, Paul R (2003), *The Political Economy of Stalinism: evidence from the Soviet archives*, Cambridge: Cambridge University Press.

Gregory, Paul R and Mark Harrison (2005), "Allocation under dictatorship: research in Stalin's archives", *Journal of Economic Literature*, XLIII (September), 721-761.

Gregory, Paul R, Philipp Schroder and Konstantin Sonin (2006), "Dictators, repression and the median citizen: an "eliminations model" of Stalin's Terror (Data from the NKVD Archives)", *CEPR Working Paper* 6014, December.

Harrison, Mark (2002), "Coercion, compliance, and the collapse of the Soviet command economy", *Economic History Review*, 55 (3), 397-433.

Ivnitskii, N. A. (2004), *Sud'ba raskulachennykh v SSSR*, Moscow: Sobranie.

Khlevnyuk, Oleg (2003), "The economy of the OGPU, NKVD, and MVD of the USSR, 1930-1953: the scale, structure and trends of development", 43-66 in: Paul R Gregory and Valery Lazarev (eds.), *The Economics of Forced Labor: the Soviet Gulag*, Stanford, CA: Hoover Institution.

Khlevnyuk, Oleg (forthcoming), *Stalin's Party Secretaries*, Stanford, CA: Hoover Press.

Kotliar, A E, ed (1983), *Zaniatost' naseleniia*, Moscow: Finansy I Statistika.

Lazarev, Valery (2003), "Conclusions", pp.189-198 in: Paul R Gregory and Valery Lazarev (eds.), *The Economics of Forced Labor: the Soviet Gulag*, Stanford, CA: Hoover Institution.

Madden, Paul (1986), *Concavity and Optimization in Microeconomics*, Oxford: Basil Blackwell.

Markevich, Andrei (2007), “The dictator’s dilemma: to punish or to assist? Plan failures and interventions under Stalin”, *Warwick Economic Research Paper* 816 (September).

Moorsteen, Richard and Raymond P Powell (1966), *The Soviet Capital Stock, 1928-1962*, Homewood, IL: Irwin.

Nove, Alec (1991) *The Economics of Feasible Socialism Revisited* London:HarperCollins

Overy, Richard (2004), *The Dictators: Hitler’s Germany and Stalin’s Russia*, London: Allen Lane.

Rogachevskaya, Lyudmila S (1973), *Likvidatsiya Bezrobotitsy v SSSR 1917-1930 gg*, Moscow: Izdatel’stvo ‘Nauka’.

Rosefielde, Stephen (1995), “Stalinism in post-communist perspective: new evidence on killings, forced labour and economic growth in the 1930s”, *Europe-Asia Studies*, 48 (6), 959-987.

Sakharov, Andrei (1975), *My country and the world*, London: Collins and Harvill

Sakharov, Andrei (1974), *Sakharov speaks*, London: Collins and Harvill

Sebag Montefiore, Simon (2007), *The Young Stalin*, London: Weidenfeld and Nicholson.

Shapiro, Carl and Joseph Stiglitz (1984), “Equilibrium unemployment as a worker discipline device” *American Economic Review*, 74 (3), 433-444.

Skidelsky, Robert (1995), *The Road from Serfdom*, London: Penguin Books.

Sokolov, Andrei (2003), "Forced labor in Soviet industry: the end of the 1930s to the mid-1950s", 23-42 in Paul R Gregory and Valery Lazarev (eds), *The Economics of Forced Labor: the Soviet Gulag*, Stanford, CA: Hoover Institution.

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

Solomon, Peter H (1996), *Soviet Criminal Justice Under Stalin*, Cambridge: Cambridge University Press.

Solzhenitsyn, Aleksandr I (1974), *The Gulag Archipelago 1918-1956: an experiment in literary investigation*, New York: Harper and Row.

Stiglitz, Joseph (1994), *Whither Socialism?*, Cambridge, MA: MIT Press.

Tikhonov, Aleksei (2003), "The end of the Gulag", pp.67-73 in: Paul R Gregory, and Valery Lazarev (eds.), *The Economics of Forced Labor: the Soviet Gulag*, Stanford, CA: Hoover Institution.

Vatlin, A (2004), *Terror raionnogo masshtaba*, Moscow: Rosspen, pp.120-215.

Wintrobe, Robert (1998), *The Political Economy of Dictatorship*, New York: Cambridge University Press.

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

Appendix 1: Custodial Population Data: sources and methods

Year	Prisons	Labour colonies	Labour camps	Total custodial population
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
1935		240,259	725,483	965,742
1936		457,088	839,406	1,296,494
1937		375,488	820,881	1,196,369
1938	548,417	336,786	996,367	1,333,153
1939	350,538	355,243	1,317,195	2,022,976
1940	190,266	315,584	1,344,408	1,850,258
1941	487,739	429,205	1,500,524	2,417,468
1942	277,992	360,447	1,415,596	2,054,035
1943	235,313	500,208	983,974	1,719,495
1944	155,213	516,225	663,594	1,335,032
1945	279,969	745,171	715,506	1,740,646
1946	261,500	956,224	600,897	1,818,621
1947	306,163	912,704	808,839	2,027,706
1948	275,850	1,091,478	1,108,057	2,475,385
1949		1,140,324	1,216,361	2,356,685
1950		1,145,051	1,416,300	2,561,351
1951		994,379	1,533,767	2,528,146
1952		793,312	1,711,202	2,504,514
1953		740,554	1,727,970	2,468,524
1959				948,000

Table A1: USSR custodial population, 1917-1953

Table A1: USSR custodial population, 1917-1953

Sources: Total custodial population 1917-1934: <http://demoscope.ru/weekly/2006/0239/tema07.php>. Prisons, colonies and camps 1934-1953: Getty *et al* (1993). Settlements: Bacon (1992). 1959: Sokolov (2003), from V.N.Zemskov, *Ukaz. Soch.*, p.15.

Notes: Total custodial population does not include those in labour settlements, as is usual in the literature. 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).

Data on the population of Soviet labour camps and colonies, labour settlements, and prisons were made available during glasnost’ from the Soviet Central State Archive. The Russian researchers who originally searched the Archive for the data were A. N. Dugin and V. N. Zemskov. Dugin’s figures were published in Western journals by Bacon (1992); these figures were checked by Zemskov, and found to be quite accurate. Zemskov’s figures were released in Getty *et al* (1993). The Archival data are not without controversy (see Ellman, 2002, for a measured discussion). Authors such as Robert Conquest (eg 1994) and Stephen Rosefield (eg 1995) have objected that the Archive figures are too low. In comparison, their own figures derived from anecdotal and personal experience of those in and around the camps would suggest that several times as many people went through the Gulag system. Nevertheless, we agree with previous arguments that the camp authorities had no incentive to run false accounts, and we also note the reported internal consistency of Archival documents (see eg Getty *et al*, 1993).

The first column shows the rather sparse data available on the numbers incarcerated in prisons, as opposed to labour camps. 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, p.1019).

Labour settlements housed kulaks – those rich peasants fortunate enough to have escaped with their lives after the forced collectivisation after 1929. Settlements were generally in remote inhospitable places, and involved (albeit relatively loosely) supervised compulsory labour related to settlement-building, such as agriculture, heavy industry and tree-felling (Overy, 2004). We will follow standard practice in excluding those in settlements from the custodial population of interest. From the point of view of labour discipline, settlements did not perform the same function as camps and colonies, in that the average worker faced no risk of being sent to a settlement. 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, p.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, p.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, p.1077) cites the case of a particular year. As Table 1 shows, 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.

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

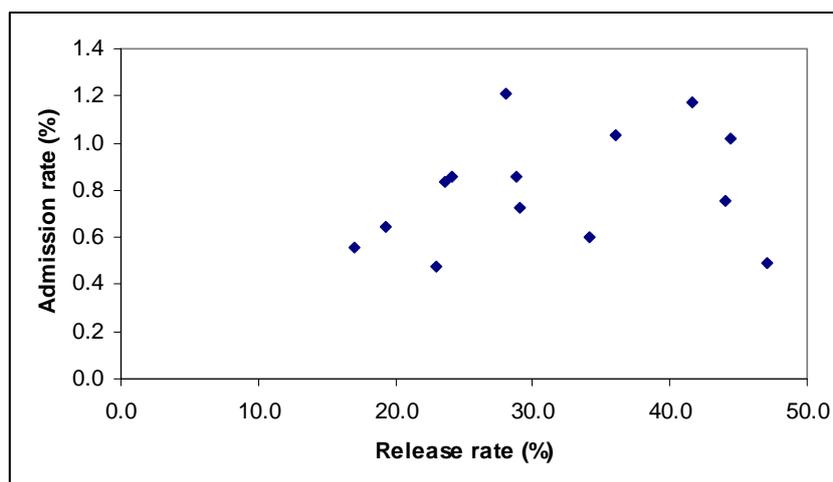


Figure A1: Release and admission rates, 1934-1947

Notes: Release rate = releases / Gulag population. Admission rate = admissions / employment.

Year	Release rate (%)	Admission rate (%)
1934	28.9	0.9
1935	29.1	0.7
1936	44.0	0.8
1937	44.4	1.0
1938	28.1	1.2
1939	17.0	0.6
1940	23.6	0.8
1941	41.6	1.2
1942	36.0	1.0
1943	34.2	0.6
1944	22.9	0.5
1945	47.1	0.5
1946	19.3	0.6
1947	24.1	0.9
1948	23.6	
1949	14.7	
1950	15.3	
1951	16.6	
1952	19.3	
1953	54.2	

Table A2: Release and admission rates, 1934-1947

Sources: Admissions: Bacon (1994). Releases: Getty *et al* (1993). Employment: Moorsteen and Powell (1966). Custodial population: See Table 1.

Notes: 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, p.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).

Year	Counter- revolutionaries	Counter- revolutionaries as % of camp population
1934	135,190	26.5
1935	118,256	16.3
1936	105,849	12.6
1937	104,826	12.8
1938	185,324	18.6
1939	454,432	34.5
1940	444,999	33.1
1941	420,293	28.0
1942	407,988	28.8
1943	345,397	35.1
1944	268,861	40.5
1945	283,351	39.6
1946	333,833	55.6
1947	427,653	52.9
1948	416,156	37.6
1949	420,696	34.6
1950	578,912	40.9
1951	475,976	31.0
1952	480,766	28.1
1953	465,256	26.9

Table A3: Political prisoners in labour camps, 1934-1953

Source: Getty *et al* (1993).

Appendix 2: Unemployment as a discipline device

The approach of Shapiro and Stiglitz (1984) is to treat a job as an asset, whose value can be enhanced by shirking but only at the risk of being fired. Consider the simplest version where being caught shirking leads to permanent unemployment. In this ‘dire punishment’ case, real income will fall from w to \bar{w} , the level of unemployment benefit, for ever. For incentive reasons the efficiency wage, w , has to be (at least) such that the saving of effort, e , by shirking matches the expected loss of welfare through becoming unemployed, i.e.

$$e = q(w - e - \bar{w})/r \quad (\text{A1})$$

where q is the hazard rate of detection and $(w - e - \bar{w})/r$ is the value of a job (capitalised at the interest rate r).

Solving for the efficiency wage with dire punishment, we find

$$w^d = e + \bar{w} + re/q \quad (\text{A2})$$

What if there is an exogenous probability of job loss, at the rate bdt , due to the flux of changing product demand, for example: how does this affect the efficiency wage? Since the job is likely to disappear anyway, its value is less. Increasing the rate of discount from r to $r + b$, valuing a job at $(w - e - \bar{w})/(r + b)$, substituting into (A1) and solving implies

$$w = e + \bar{w} + (r + b)e/q \quad (\text{A3})$$

So random break-ups increase the efficiency wage.

Such random inflows into unemployment will, in steady state, need to be matched by outflows.

The authors assume that unemployment is temporary, with access to jobs from the state of unemployment at the rate of adt ; and the unemployed are effectively anonymous with no stigma attached to having been fired for shirking. The effect of incorporating re-entry to employment is to further increase the rate of discount on the RHS of equation (A1) so that the efficiency wage becomes

$$w = e + \bar{w} + (r + a + b)e/q \quad (\text{A4})$$

The dynamics of unemployment are such that unemployment will increase if the number of break-ups bL exceeds the number of jobs obtained $a(N - L)$, since

$$\dot{u} = bL - a(N - L).$$

In steady state equilibrium where inflows into unemployment match outflows, the rate of job access and break-up must satisfy the condition that

$$au = b(1 - u)$$

where u denotes the unemployment rate.

For given values of b and a , unemployment would be increasing to the right of L in Figure A1 and decreasing to the left of L . In deriving the NSC, however, SS assume that a is endogenous and will adjust to support any given b . This means it has to rise without limit to enable stationary states with very low unemployment. The rate of job acquisition will be very rapid at low rates of unemployment. As this means that the punishment involved in unemployment is vanishingly small, the NSC goes off to infinity when unemployment is low.

Since this implies

$a + b = b/u$, the efficiency wage they derive for capitalism is

$$w = e + \bar{w} + (r + b/u)e/q \quad (\text{A5})$$

This has the property that the efficiency wage goes to infinity as u falls to zero: the access rate has to increase sharply to satisfy the equilibrium conditions just described, so unemployment becomes vanishingly transitory.

Figure 1 shows the NSC curve, along with marginal product of labour (MPL) curve and the equilibrium NSC=MPL condition.⁵¹ The Figure also shows the quit rate b and job acquisition rate a ; at the stationary equilibrium, $au = (1 - u)b$. In the case of dire punishment, unemployment is permanent, so $a = 0$.

Note that, if the incentive conditions are satisfied, the pool of unemployed acts as a credible threat. In equilibrium there are no shirkers among the unemployed.

⁵¹ The level of output is the area under the MPL curve to the left of B.

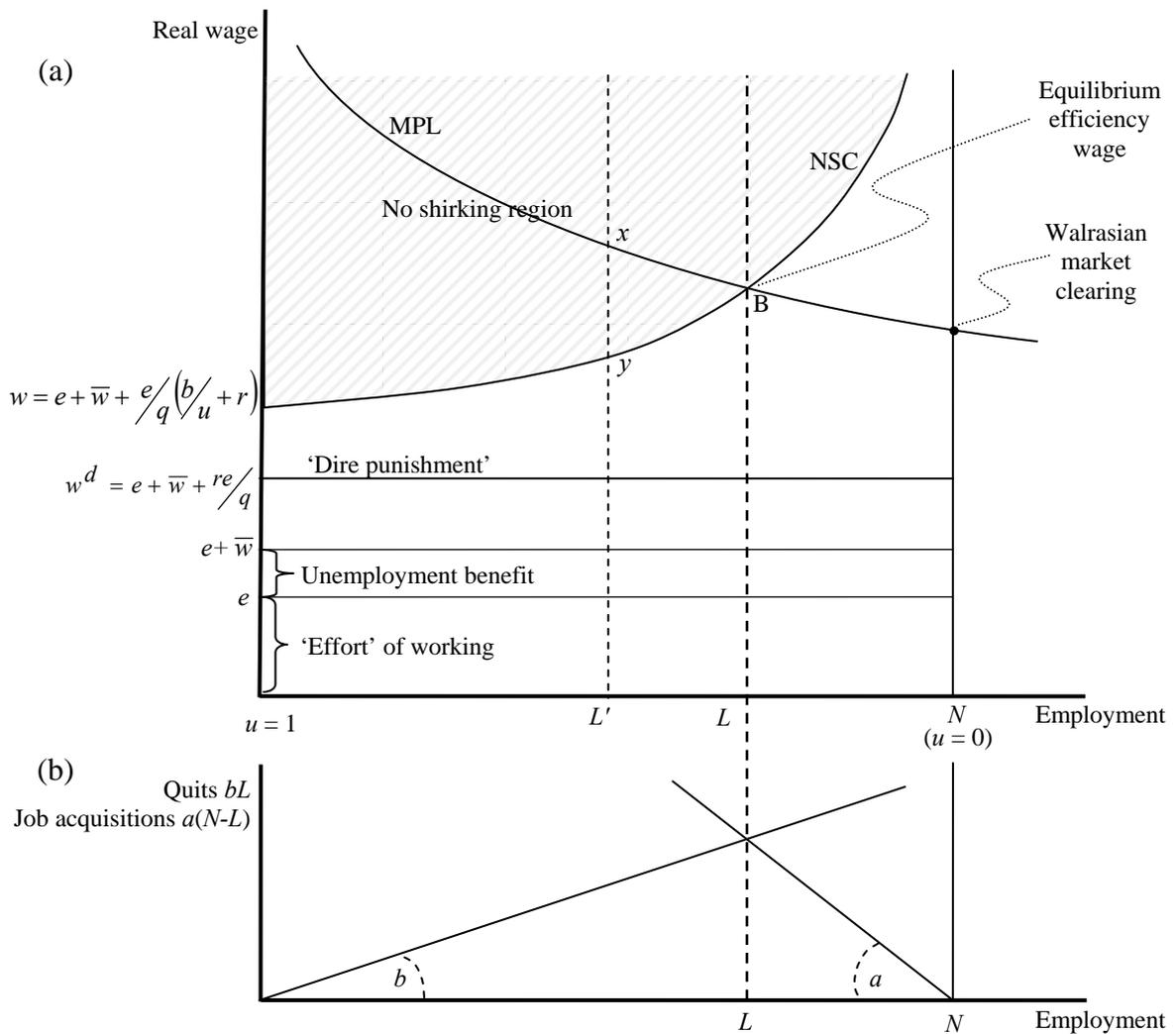


Figure A2: The Shapiro-Stiglitz model, including 'dire punishment'

This model is implicitly developed for an economy not suffering from demand failure (so it is reasonable to talk about high-employment equilibria on the MPL curve). But in the 1930s, at the time when the USSR had eliminated unemployment, western free-market economies were suffering from mass unemployment and substantial disequilibrium in the labour market. One can appeal to the logic of 'quantity-constrained' economics to show this in Figure 1. Let employment be at L' to the left of L , with the decline measured by the integral of MPL between L and L' , and the real wage lying somewhere between x and y , depending on the bargaining strength of employers and workers. Such a non-market-clearing equilibrium might be a better representation of the state of western economies at the time that Stalin's experiment in coercion began.

Appendix 3: Derivation of Propositions

Proposition 1

In steady state equilibrium, with the prison population constant, the flow of random incarcerations must match the flow of those being released and dying in prison, so

$$\pi(1-p) = (\rho + \delta)p$$

where p denotes the percentage of the workforce in prison and δdt the death rate.

$$\text{Hence } \rho = -\delta + \pi(1-p)/p$$

$$\text{and so } \rho + \pi = -\delta + \pi/p .$$

By treating ρ as endogenous (subject to a maximum, $\bar{\rho}$), one may define a multiplicity of steady state equilibria. Substituting for $\rho + \pi$ in the equation defining the efficiency wage, namely:

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

one obtains the result

$$w^*(p) = e + \gamma + (r - \delta + \pi/p)e/q$$

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

Proposition 2

The dictator's problem can be written as

$$\max_L f(F(L) - wL) \text{ subject to } w - w^*(L) \geq 0$$

where the objective is to maximise investible funds $F(L) - wL$, and the constraint is that wages satisfy the efficiency condition $w \geq w^*(L)$, where $w^*(L)$ the efficiency wage.

Formally, the Lagrangian for this problem can be written as

$$\Lambda(L, \lambda; \pi, \gamma) = F(L) - wL + \lambda(w - w^*(L)) + s(N - L)$$

The First Order Conditions are

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

$$\partial\Lambda / \partial L = F'_L - w - \lambda w'_L - s = 0$$

Since the objective is non-stationary in w , the Lagrange multiplier is positive ($\lambda = L > 0$)

and $w = w^*(L)$; hence the FOC can be written:

$$F'_L = w^* - Lw'_L - s$$

Proposition 3

As noted in deriving Proposition 2, the Lagrangian for this problem is

$$\Lambda(L, \lambda; \pi, \gamma) = F(L) - wL + \lambda(w - w^*(L))$$

where the Lagrange multiplier $\lambda = L$ is the size of the civilian labour force. But the multiplier measures the rate of change of the optimal value of investible resources (the dictator's objective) with respect to perturbations which ease the incentive constraint, i.e. with respect to alterations of parameters that affect the efficiency wage. (See Madden (1986) for further discussion of the 'envelope theorem' being used here.)

But with the efficiency wage determined as $w^*(L) = e + \gamma + (r - \delta + \pi / p)e / q$ and

$\partial w^* / \partial \gamma = 1$, increases in γ will tighten the incentive constraint and lower the objective, i.e.

$$\partial \Lambda / \partial \gamma = (\partial \Lambda / \partial w^*)(\partial w^* / \partial \gamma) = -\lambda < 0.$$

Per contra the objective will increase if γ is reduced, i.e. investible resources will be decreasing in living standards in the Gulag.

Proposition 4

This is another application of the envelope theorem. As noted in discussing Proposition 3, the Lagrangian for this problem is decreasing in the efficiency wage which is determined as

$w^*(L) = e + \gamma + (r - \delta + \pi / p)e / q$. Since $\partial w^* / \partial \pi > 0$, investible resources will be decreasing

in the extent of randomised terror, i.e. $\partial \Lambda / \partial \pi = (\partial \Lambda / \partial w^*)(\partial w^* / \partial \pi) < 0$.