Chapter 1

The Other Government: An Introduction to the Missing Facts

Can a country be less productive after Communism than under Communism? This question was unthinkable ten years ago and is unspeakable today. The past decade of Russian experience raises such a possibility. During 1990-99, Russia lost 45 percent of its gross domestic product (GDP) and about 50 percent of consumption.[1] This dwarfs the experience of the U.S. Great Depression when output declined by 30 percent and consumption fell by about 20 percent. Even during the extreme social experiment in the Soviet Union in the late 1920s-early 1930s, with its severe demographic losses, the economic decline was smaller than in the 1990s. In addition to protracted contraction, Russia underwent serial defaults during 1992-99. They peaked in August 1998 with what one may call the Great Default, when the government repudiated its domestic bonds. After that, periodic defaults on external debt occurred in 1999.[2]

Economic contraction of the 1990s was protracted and continuous. After a steep output decline during 1992-94, contraction slowed down in 1995-96 and reversed into a slight recovery in 1997, only to revert to another big slide in 1998. Part of the 1998 drop in GDP was recovered in 1999 and early 2000, but a pronounced decline in personal consumption continued through 1999. Many other post-Communist economies experienced great contractions. Their causes remain elusive after a decade of study.[3]

---

[1]GDP: The International Monetary Fund, *World Economic Outlook*, October 1999 (Washington, D.C.: The IMF, 1999), p. 126, and *World Economic Outlook*, May 2000, pp. 27, 113. Consumption: Russian State Committee on Statistics (RSCS), various issues. Independent computations confirm the magnitudes found in the official data. See McKinsey Global Institute, "Unlocking Economic Growth in Russia" (Moscow: McKinsey & Co., 1999), available at http://205.253.128.117/mgi/russian.htm; and, Mark De Broeck and Vincent Koen, "The Great Contractions in Russia, the Baltics and the Other Countries of the Former Soviet Union: A View from the Supply Side," IMF Working Paper WP/00/32 (March 2000).

[2]From December 1998 to June 1999, Russia missed $1.2 billion in payments on the previously restructured dollar-denominated bonds. In June 1999 and in early December 1999, it failed to pay $963 million to the London Club of commercial creditors, and missed another $550 billion payment in late December of that year. A series of multi-billion dollar reschedules with the Paris Club of government creditors and the London Club of commercial creditors in 1999 and 2000 saved them and Russia from a wholesale default on the total external debt.

[3]For an excellent evaluation of the causes of the Great Contraction across many post-Communist economies see Robert A. Mundell, "The Great Contractions in Transition Economies," in Mario I. Blejer and Marko Skreb, eds., *Macroeconomic Stabilization in Transition Economies* (Cambridge, England, and New York: Cambridge University Press, 1997), pp. 73-99. The term "Great Contraction" comes from this article and is now widely used. See Mark De Broeck and Vincent Koen, "The Great Contractions in Russia,..." The most extensive statistical investigation to-date attributes contraction in many post-Communist countries to the decrease of subsidized trade with Russia. See Andrew Berg, Eduardo Borensztein, Ranta Sahay, and Jeromin Zettelmeyer, "The Evolution of Output in Transition Economies:

The range of possible explanations and remedies relies on the experience of developing countries.[4] If this approach was sufficient, Russia's Great Contraction and the Great Default could have been averted or at least predicted. At the very least, they would have been explained. Instead of looking for what is familiarly wrong, we will search for what is wrongly unfamiliar. This chapter introduces the facts missing in the conventional approach and lays out the groundwork for our book. It outlines how the network of enterprises, inherited and liberated from central planning, has taken over fiscal and monetary power. The enterprise network has become a parallel government. It redistributes all incomes in the economy and causes defaults and protracted contraction. The next chapters add detailed analysis and evidence.

The Missing Facts and the Key Empirical Relations

What is missing? What is missing is a number of core facts and an understanding of the core structure of the new Russian economy.

1. The Fake Budget

To start with, Russian fiscal accounts, the centerpiece of analysis and policy guidance by the Russian government, U.S. Treasury, the IMF, Western investors, and many Western researchers, are an obvious fake. Any apprentice accountant can tell this in a minute. To cite one example of many, tax offsets and Central Bank credit to enterprises for paying taxes should be added to budget expenditures or subtracted from tax revenues.[5] Either way, they should be added to the budget deficit.[6]

Let us see how the Russian budget works. Suppose that one or more enterprises provide some

---

Explaining the Differences," IMF Working Paper WP/99/73 (July 1999). Even if this explanation was valid (we will show in Chapters 14 and 15 why it is not), it cannot apply to Russia and thus cannot be universal.

[4]For two overviews, see Dani Rodrik, "Understanding Economic Policy Reform," *The Journal of Economic Literature* 34, no. 1 (March 1996): 9-41 and Mathias Dewatripont and Gerard Roland, "Transition as a Process of Large-Scale Institutional Change," *Economics of Transition* 4, no. 1 (May 1996): 1-30. A most influential compendium of this approach is The World Bank, *From Plan to Market. World Development Report 1996* (New York: Oxford University Press for the World Bank, 1996). The latest summary is in Stanley Fischer and Ratna Sahay, "The Transition Economies After Ten Years", IMF Working Paper WP/00/30 (March 2000).

[5]Tax offsets are tax credits. The government allows enterprises to reduce their tax liability in the amount of the value of goods and services they rendered to government agencies, the military, non-profit organizations, and key industries. For a detailed description and data, see Brian Pinto et. al., "Dismantling Russia's Nonpayments System" (Moscow: The World Bank, September 1999). Separately, the Central Bank prints money and disburses it through the banking system as loans to enterprises, after which enterprises remit taxes to the government.

[6]The IMF lists several factors which affect the accuracy of reported budget deficits in post-Communist economies, but none of the above. See Stanley Fischer and Ratna Sahay, "The Transition Economies After Ten Years," p. 8.

goods or services to government organizations but don't receive payment. Or, perhaps an enterprise may be mandated by the government to supply another enterprise and, again, doesn't get paid. In exchange for delivery of goods and services, but for which payment was not received, the government forgives (offsets) the enterprises in question a commensurate amount of tax payments. The forgiveness of a tax liability, or a tax offset, enables enterprise to receive what they are due by retaining for themselves the equivalent amount in taxes they withheld from workers and collected from consumers, which otherwise they would (should) remit to the government in cash. These payments by government, tax offsets, represent government expenditure which is not recorded in the budget.

Suppose the government only counts cash flow as it treats the budget. Since the tax offset, the equivalent of government expenditure, was not paid in cash, there is no need to record it in a purely cash flow budget. But the tax revenues collected by the government were only partly received in cash, inasmuch as the offset amount was not received in cash. But the government lists the offset as part of revenues. The result is that either expenditures are understated if full accounting is made or revenues overstated in the case of cash flow accounting.

The same reasoning applies to Central Bank credit to enterprises through the banking system. Enterprises are given credits with which to pay taxes. If they did not receive this credit and did not pay the commensurate amount in taxes, recorded government revenues would have been smaller by this very amount. If enterprises were given a direct budget subsidy instead of Central Bank credit, recorded government expenditures would have been greater by this amount. To cover the budget deficit resulting from lower revenues or higher expenditures, the Central Bank could purchase government bonds and print money in the same amount, instead of giving credit to enterprises. Again, either expenditures are understated or revenues overstated.

Due to these and similar factors, the true budget deficit in Russia was 10 to 15 percentage points of GDP higher in various years than universally accepted. One can readily infer that Russia's fiscal house is unsustainable and that Russia must default every few years, which has been the pattern.

2. The Tax Subsidy

The next key fact follows directly. Russia runs a huge, off-budget subsidy to enterprises in the amount of 15-25 percent of annual GDP in various years.[7] This subsidy is even more strange than it is huge. It is not like any of the variety of subsidies that we know through history and around the world. What unites all subsidies existing on record is that the government gives them, voluntarily or under pressure. Subsidies emanate from the government. In contrast, the subsidy in Russia is self-made. Enterprises collect taxes from consumers and workers on behalf of the government, as do firms in market economies, but do

[7]Chapter 9 presents detailed data on the bulk of this subsidy but cannot capture it in full.

not remit part of the revenue they collect to the government.[8] They also receive Central Bank credit to increase remittance, which is a reshuffle of the same subsidy.[9] This subsidy is directly taken by enterprises from the public, after which part of the take is exchanged with the government for Central Bank credit. Therefore, even the Central Bank part of the subsidy is not given by the government. The entire subsidy is not given, it is taken. It does not emanate from the government. It is collected as a tax and should be called the tax subsidy. Its simple essence is parallel taxation of the public by enterprises.

The network of enterprises acts as a parallel government, collecting revenues through the interplay of taxing the public and printing money. It follows that this second government co-owns the tax base and the Central Bank. The enterprise network is the other fiscal authority and has power over the monetary authority. Since the tax subsidy is off-budget and is self-collectible, and since enterprises decide how much of tax collection they give to the government, their subsidy determines the ultimate expenditure. They decide on the size of the budget deficit and debt, and they are the ultimate fiscal authority. The G-7 finance ministers and central bank governors are talking with the penultimate authority at meetings with their Russian counterparts.[10]

3. The Tax Subsidy is a New Phenomenon

The tax subsidy may look like some familiar off-budget subsidies. It may resemble bailouts and tax exemptions in Western market and developing economies and the soft budget constraint in central plan economies. The fundamental difference of the tax subsidy is that all other subsidies emanate from the government, whereas the tax subsidy is self-collected. In Western market and developing economies, firms

---

[8]The literature described part of this non-remittance under the name of tax arrears. See Mark E. Schaffer, "Government Subsidies to Enterprises in Central and Eastern Europe: Budgetary Subsidies and Tax Arrears," in David M.G. Newbery, ed., *Tax and Benefit Reform in Central and Eastern Europe* (London: Center for Economic Policy Research, 1995), pp. 115-144; Mark E. Schaffer, "Do Firms in Transition Economies Have Soft Budget Constraints? A Reconsideration of Concepts and Evidence," *Journal of Comparative Economics* 26, no. 1 (March 1998): 80-103; Michael S. Bernstam and Alvin Rabushka, *Fixing Russia's Banks* (Stanford, CA: Hoover Institution Press, 1998), pp. 29-34, 69-71; and, Brian Pinto et. al, "Dismantling Russia's Nonpayments System."

[9]The literature keenly calls the latter phenomenon a "quasi-fiscal subsidy." See G.A. (Sandy) Mackenzie, "The Hidden Government Deficit," *Finance and Development* 31, no. 4 (December 1994): 32-35 and The World Bank, *From Plan to Market*, pp. 35-36.

[10]To borrow a term from biology, one can call the relationship between enterprises and the government symbiotic, whereby the government is the symbiont. The term enterprise is a shortcut. James M. Buchanan taught all of us methodological individualism. See James M. Buchanan, *The Limits of Liberty: Between Anarchy and Leviathan* (Chicago: The University of Chicago Press, 1975). When we talk about enterprises, we mean their individual decision makers, such as owners and managers and, whenever applicable, workers. For clarity, we have chosen to use the term enterprise to denote productive units (even if money losing) in Communist and post-Communist economies, as opposed to market firms in Western market economies. We can also use the terms businesses and companies to describe market firms. When we need a generic term for both firms and enterprises, we will call them establishments.

obtain subsidies by lobbying the government and manipulating the political process[11] or receive bailouts because they are "too big to fail."[12] The best known bailouts are those of the Chrysler Corporation in 1979, the Savings and Loan Associations in the 1980s, and numerous corporate and banking bailouts in Japan and other countries.

An example of the off-budget subsidy under central planning is the cash flow subsidy, known in the literature as the soft budget constraint.[13] This is a peculiar subsidy. On first glance, it looks like a bailout in Western market and developing economies: The government automatically picks up the negative money balances of enterprises when they run into cash flow problems and cannot meet their payment obligations. Beneath the surface, however, this cash flow subsidy is a forced subsidy. It is imposed by the government on enterprises in order to enforce such key features of central planning as the forced production mix and mandated deliveries to designated buyers—in short, forced exchange. Without this cash flow subsidy to the buyers, the sellers would break free, reorient supplies to paying customers, and change their production mix, as they always try to do. This would destroy central planning. The soft budget constraint is not a subsidy bonanza, as it is portrayed in the literature; it is the ultimate enforcement of central planning. It is a coordination mechanism of central planning, a chain that ties together the gang production on the unified assembly line throughout the economy.

The tax subsidy in post-Communist economies is so unprecedented and incredible that the literature usually confuses it with the soft budget constraint.[14] In reality, they are exact opposites. The soft budget constraint is the subsidy that the government wants and enterprises do not. The tax subsidy is the subsidy that enterprises want and the government does not. The tax subsidy is self-taken, collected by the enterprise network as the ultimate fiscal autohority.

4. The Takeover of the Tax Base

---

[11]The literature amply calls this phenomenon rent-seeking. The founding works on rent-seeking are Gordon Tullock, "The Welfare Costs of Tariffs, Monopolies and Theft," *Western Economic Journal* 5, no. 2 (June 1967): 224-232; Anne O. Krueger, "The Political Economy of the Rent-Seeking Society," *American Economic Review* 64, no. 3 (June 1974): 291-303; and, James M. Buchanan, Robert D. Tollison, and Gordon Tullock, eds., *Toward a Theory of the Rent-Seeking Society* (College Station: Texas A&M University, 1980). For recent developments see Robert D. Tollison and Roger D. Congleton, eds., *The Economic Analysis of Rent-Seeking* (Aldershot, England and Brookfield, Vermont: E. Elgar, 1995).

[12]George A. Akerlof and Paul M. Romer, "Looting: The Economic Underworld of Bankruptcy for Profit," *Brookings Papers on Economic Activity* no. 2 (1993), pp. 1-73.

[13]Janos Kornai, "The Place of the Soft Budget Constraint Syndrome in Economic Theory," *Journal of Comparative Economics* 26, no. 1 (March 1998): 11-17.

[14]E.g., The World Bank, *From Plan to Market*, p. 142; Mark E. Schaffer, "Do Firms in Transition Economies Have Soft Budget Constraints?"; and, Brian Pinto et. al., "Dismantling Russia's Nonpayments System."

What empowers enterprises? What enables them to exact the tax subsidy? The answer may, indeed should, sound incredible: Russian enterprises can bloc the flow of payments, shut down the economy, and deprive the government of the tax base. The flow of payments holds together any economy that uses money; the flow of tax revenues holds the government at power. If enterprises can create a payment crisis, they can seize the tax collection with impunity. The case of a payment crisis is exceptionally rare; the case of the takeover of the tax base borders on the impossible. Nevertheless this is the case which we find in Russia and similar countries.

Imagine if enterprises halt payments to each other for inputs. They simply stop paying bills, discard invoices, terminate supplies to non-paying customers, and switch to barter in the rest of the transactions. This would wipe out the tax base, which consists of payroll, personal income, and sales, all embodied in monied payments. What can the government do in this situation? Its options range from bad to worse. The government may try to borrow funds. But who would lend to a government that has lost its tax base and is unable to repay debt? The government can print money, but it would quickly run into hyperinflation in the absence of other sources of financing (taxes and debt). People would switch to dollars, abandon local currency altogether, and the government could not buy pencils and hire soldiers for rubles. No tax revenues, no borrowing, not much of a money printing option. The government would go out of commission.

Being on the brink of losing the tax base on the backdrop of a payment crisis is not a fantastic possibility. This has been the daily existence of the Russian government. Russia has been chronically on the edge of this scenario since January 1992 when central planning was abolished and enterprise transactions were decontrolled. Under central planning, the government helped enterprises pay their bills. Enterprises, however, were constrained in the amount of their invoices and bills by price controls which were linked to production targets. Once central planning was abolished and transactions and prices decontrolled, enterprises were free to charge each other and the government whatever amount of invoices they thought—perhaps initially hoped—would be paid. Either by sheer accident or design, a number of enterprise managers started to inflate their invoices as claims on public income to test if they would be paid. Once the results were known, everyone followed suit. Those who stayed on the sidelines lost out. By attrition, the remaining players were, through a new invisible hand, engaged in a seemingly coordinated raid on the public income through the mechanism of over-invoicing to get paid as much as they could.

With overcharging as the norm, enterprises create a traffic jam of payments throughout the economy, whereby it takes a buyer three-to-four months to pay an invoice to the seller, after which he can pay his own suppliers.[15] A further increase in this jam of payments can grind the economy to a halt.

---

[15]Calculated from the data from the Russian State Committee on Statistics. Throughout this book, all data derives from the Russian State Committee on Statistics and the Central Bank of Russia, unless we specifically cite other sources. Most of this data is in the public domain. It appears in various printed annual, quarterly, monthly, and occasional publications of these two agencies. Much of this data is available on the Internet, in both English and Russian, at the following sites: http://www.cbr.ru/, http://www.gks.ru/, http://www.rbs.ru/gks/, and

Payments in the economy are like the air we breathe; a prolonged interruption is lethal. Payments embody exchange. Without payments, there is no exchange save small-time barter. Payments are the cash flow which lubricates the production flow. The traffic jam of payments is a cash flow jam. But the cash flow is fungible like liquid in pipes. As Gertrude Stein summed it up, "[m]oney is always there but the pockets change." This is exactly the point. In the payment jam, on the edge of a complete stoppage of the traffic of cash flow, production flow, and tax flow, the pockets become fungible. Enterprises take the cash flow as needed from whatever is there. What is there is the taxes they have collected from the public for the government. Enterprises appropriate the cash flow from the tax base, pay their accumulating bills, keep the traffic of payments going, and thus preserve the rest of the tax base for the government.

5. In the Payment Jam

What is a traffic jam? People and horses knew it in the streets of London in the late 18th century, but hardly anywhere else. Now people know it in every corner of earth, with automobiles, rickshaws, and what not. The development of the information economy and home offices will retire this knowledge. Fifty years from now, if not earlier, only old folks will remember in the West what a traffic jam is, and they might not be able to relate it to others. One needs to be in it to know it. The same with the payment jam. Try telling it to those who have not been in it.

The traffic jam of payments is one of the most important, most confusing, and most ignored facts of life in Russia. This payment jam is peculiar to economies after central planning.[16] All post-Communist economies experienced it to one or another extent, for a longer or shorter period. A small body of peripheral literature extensively documented and analyzed this phenomenon but did not explain its power. The literature calls this phenomenon a liquidity crisis (in the case of Yugoslavia of the 1960s and 1970s), non-payments (in the case of Russia), and arrears.[17] The preponderant literature has long maintained that

---

http://www.finmarket.ru/meconomic.asp/. In addition, the Russian-European Center for Economic Policy collects from Russian official agencies and provides to the public an extensive, ten-year data set and monthly updates at http://www.hhs.se/site/ret/ret.htm and http://www.hhs.se/site/ret/exceldb/default.htm.

[16]Germany after the War Socialism of World War I and the Soviet Union in the 1920s-early 1930s, between the two episodes of central planning, saw this phenomenon. See Gerald D. Feldman, *The Great Disorder: Politics, Economics and Society in the German Inflation, 1914-1924* (New York and Oxford, U.K.: Oxford University Press, 1993); and, Paul Gregory and Aleksei Tikhonov, "Money, Credit, and Plan: Creating the Soviet Financial System," Hoover Institution, Working Papers in International Studies, I-99-7 (October 1999).

[17]Laura D. Tyson, "Liquidity Crises in the Yugoslav Economy: An Alternative to Bankruptcy?" *Soviet Studies* 29, no. 2 (April 1977): 284-295; P.T. Knight, "Financial Discipline and Structural Adjustment in Yugoslavia: Rehabilitation and Bankruptcy of Loss-Making Enterprises," World Bank Staff Working Papers, no. 705 (Washington, D.C.: The World Bank, 1984); Guillermo A. Calvo and Fabrizio Coricelli, "Credit Market Imperfections and Output Response in Previously Centrally Planned Economies," in Gerard Caprio, David Folkerts-Landau, and Timothy D. Lane, eds., *Building Sound Finance in Emerging Market Economies* (Washington, D.C.: The International Monetary Fund and the World Bank, 1994), pp. 257-294; Guillermo A. Calvo and Fabrizio Coricelli. "Inter-Enterprise Arrears in Economies in Transition," in Robert Holzmann, Janos Gacs, and George Winckler, eds., *Output Decline in Eastern Europe: Unavoidable, External*

arrears do not matter and are bound to dissipate over the course of financial development.[18] An influential book argued that the issue of arrears is a fallacy because late payments between firms are a normal problem in Western market economies,[19] which is patently untrue and structurally impossible.

The average length of payments in market economies does not exceed a certain designated time thanks to a simple self-regulating mechanism: Sellers do not ship goods to delinquent payers; buyers slow down purchase orders when they run into cash flow problems. The payment jam does not and cannot exist in market economies. It also does not and cannot exist in central plan economies: The government automatically provides liquidity to replenish enterprises cash flow shortfalls in order to enforce production and exchange according to government plan. This is the essence of the soft budget constraint.

The issue of the payment jam is hard to grasp. The subject matter escapes words and statistical pinpointing; we will leave it nameless for the time being and christen it later. Arrears, non-payments, and similar terms blind the view. They imply the stock of debt on a given date, for example, U.S. federal income taxes for 1999 fell into arrears after April 17, 2000. But while each individual invoice or bill has a date due, the flow of payments does not (unlike income taxes). Every day, goods are shipped, invoices are sent and become receivables, bills arrive and become payables (one man's invoice is another man's bill; one man's receivable is another man's payable), and payments are made—money comes, money goes, money changes pockets, the cash flows.[20] This process of shipping goods with invoices and collecting

---

Influence or Homemade? (Dodrecht, Boston, and London: Kluwer Academic Publishers, 1995), pp. 193-212; Enrico C. Perotti. " A Taxonomy of Post-Socialist Financial Systems: Decentralized Enforcement and the Creation of Inside Money," *Economics of Transition* 2, no. 1 (January 1994): 71-81; Enrico C. Perotti, "Inertial Credit and Opportunistic Arrears in Transition," *European Economic Review* 42, no. 9 (November 1998): 1703-25; Fabrizio Coricelli, *Macroeconomic Policies and the Development of Markets in Transition Economies* (Budapest: Central European University Press, 1998), pp. 52-85; Clifford G. Gaddy and Barry W. Ickes, "Russia's Virtual Economy," *Foreign Affairs* 77, no. 5 (September-October 1998): 53-67; World Bank, Europe and Central Asia Region, Energy Sector Unit, "Non-Payment in the Electricity Sector in Eastern Europe and the Former Soviet Union," Technical Paper 423 (Washington, D.C.: The World Bank, 1999); Brian Pinto et. al., "Dismantling Russia's Nonpayments System"; Michael S. Bernstam, "A Proposal for Solving the Financial Crisis of the Spring and Summer of 1992," Russian Government's Center for Economic Reform, *Information Bulletin*, no. 4 (October 1992): 2-5; Michael S. Bernstam and Thomas E. MaCurdy, *Inter-Enterprise Debt and the Russian Coal Industry, 1992-94* (Washington, D.C.: Partners in Economic Reform, Inc., for the USAID, 1996); and, Michael S. Bernstam and Alvin Rabushka, *Fixing Russia's Banks* (Stanford: Hoover Press, 1998), pp. 28-33, 69-71, 84.

[18]The IMF, *World Economic Outlook*, May 1996 (Washington, D.C.: The IMF, 1996), pp. 88-90. The IMF reversed its position after the Great Default and the collapse of financial stabilization, but has offered no analysis. See The IMF, *World Economic Outlook*, October 1999, p. 70.

[19]Richard G. Layard and John Parker, *The Coming Russian Boom: A Guide to New Markets and Politics* (New York: The Free Press, 1996), p. 133ff.

[20]Receivables and payables, or accounts receivable and accounts payable, are accounting terms denoting invoices and bills, respectively. They all have a due date, from 10 to 90 days within which to be paid, but most invoices give 30 days to pay. Late payments entail financial charges; early payments (e.g., within 10 days on a 30-day invoice)

payments thereafter is known as trade credit; it represents one of the greatest inventions in human history and has developed in modern economies since the bill of exchange emerged in the Levant in the eighth century and in Europe in the thirteenth century. What matters is not how much of the stock of total payments is in arrears (overdue), but whether the flow of payments can go on smoothly, without financial losses, production frictions, or a failure to meet fiscal obligations. This may look like a subtle difference with the stock of overdue invoices, but it makes a not so subtle difference between a normal flow of payments and a payment jam, a cash flow crisis.[21]

A true indication of a payment jam is not the share of arrears but the average length before payments are made. If this length is beyond the limits set by sellers for a smooth flow of payments, the entire outstanding stock of receivables in the economy is in arrears. Not some but all bills are in arrears in the flow sense. Once we recognize that the flow, not the stock, matters, it is a matter of the weighted average of the age of outstanding invoices. Consider a numerical example. Bills may be due within 10, 20, 30, 60 or more days; most invoices request payment within 30 days. In the U.S., the outstanding stock of receivables constitutes $1.4 trillion, or about 15 percent of GDP.[22] Suppose $0.84 trillion, or 60 percent of the total, were paid on the 40th day. A gigantic sum and an impressive statistic: 60 percent of bills in arrears, more than in Russia, whereby 50 percent of the stock of receivables is in arrears. Suppose the other 40 percent of the U.S. total were paid on the 10th day. The average length of payment is 28 days, that is, although 60 percent of the stock of outstanding bills is in arrears, the entire flow of payments is not. Every day some payments arrive. The average firm and all firms command the same cash flow, the same money balances in the bank as if all their invoices were paid on the 28th day. On the average day, firms receive $50 billion in payments and can manage their cash flow, sales, production, payroll, and tax liabilities smoothly. In reality, U.S. receivables clear around 30 days on the average, but the distribution is more even than in our example. In Russia, in contrast with the U.S., the average length of payments is between three and four months, and the entire stock of receivables is therefore in arrears.[23] Like in the traffic jam,

---

often carry a 1 to 2 percent discount, implying an 18.25 percent to 36.5 percent annualized interest rate. This shows the cost of the smooth cash flow for the seller.

[21]A cash flow crisis is not a liquidity crisis. The latter notion is a misnomer in the context of profligate money printing and high inflation. It is not that the cash pool is too small but rather that the invoice is too big. We discuss this shortly.

[22]Board of Governors of the Federal Reserve System, Flow of Funds Accounts, available at http://www.bog.frb.fed.us/releases/Z1/Current/z1.pdf.

[23]This simple statistical point is crucial for measurement of overdue invoices, especially in empirical tests of their impact on economic activity. The literature misses this point and the fact that all receivables in Russia and similar countries are in arrears. The literature uses the meaningless data on the stock of overdue receivables and payables, which measures only an unweighted and spurious portion of the outstanding stock. See Fabrizio Coricelli, *Macroeconomic Policies*, pp. 71-85; and, Brian Pinto et. al., "Dismantling Russia's Nonpayments System." Regrettably, the most extensive database on the Russian economy, compiled by the Russian-European Center for Economic Policy under the auspices of the European Community and the Stockholm School of Economics, offers only the same meaningless series on the stock of overdue bills. See http://www.hhs.se/site/ret/ret.htm and

either all vehicles are in it on a given road, or none.

But even that is not sufficient to create a payment crisis. Average payments may be late, all invoices may be in arrears, but the cash flow may suffice. The traffic of payments may move slowly but smoothly. The real issue is not late payments, arrears, not even their overdue length. The critical issue is whether enterprises can pay their outstanding bills from their own cash flow, without taking over (and from) the tax base. If they can, there is no payment jam. Then the government can enforce full tax remittance. The government can stop enterprises from taking the tax subsidy. But if invoices due on the average day exceed the average daily cash flow of enterprises—if invoices are excessive and fall into prolonged arrears—enterprises face two options. They can halt payments to sellers and cut supplies to delinquent buyers. This is the first option. Alternatively, they can take cash from the tax collection, that is, take the tax subsidy. This is the second option. In this case, the government loses part of tax remittance. But if payments stop, shipments stop, except for barter. This constricts production flow and halts the entire tax flow. The government loses the entire tax base. Given this alternative, the tax subsidy enforces itself automatically.

The situation on the edge of halting the payment flow between enterprises, whereby enterprises use tax payments to make their payments to suppliers, characterizes the payment jam. Its simple criterion is that enterprise pockets and the government pocket become automatically fungible. Not only payments are fungible, like everywhere else, but also the payees—the government and enterprises—are fungible. Instead of taxing the public to pay the government, enterprises tax the public to pay themselves.

The critical issue is not overdue invoices. It is **excess invoices**—excess in relation to the cash flow before the tax subsidy is collected. They, and not arrears, create the payment jam. Excess invoices are simply enterprise claims on the tax subsidy, claims on public income, which are self-enforceable on the edge of payment and production breakdown.

6. The Fiscal System the World Never Saw

Each period of time that enterprises help themselves to government tax revenues, they replenish their cash flow and meet their payments to suppliers. Let us freeze the picture after any period of paying off excess invoices and then unfreeze it. From this point on, enterprises command sufficient cash flow to roll over the existing volume of receivables during their payment period. The sellers receive payments and ship new supplies with new invoices; the buyers pay their bills, place new orders, and obtain new inputs. The period of payment may still be long by Western standards, and even longer than before, and bills may still be in arrears, but the economy is off the edge of halting payments, production, and tax flows. The payment jam is over. But then so is the tax subsidy. Without the payment jam, without the economy on the brink of breaking down, enterprises cannot enforce the tax subsidy. In other words, the government

---

http://www.hhs.se/site/ret/exceldb/default.htm. We provide the relevant data on the total outstanding stock of receivables in arrears in the statistical appendix.

can enforce full tax remittance.

The enterprise network has to create incremental invoices, necessarily in excess of enterprise cash flow, in order to perpetuate the tax subsidy. Enterprises need to place the economy again in the payment jam, on the edge of halting production and tax flows. Then they collect the tax subsidy again. The economy goes on and off the edge of halting payments in a perpetual circle. Enterprises are free to charge the public with excess invoices and collect the tax subsidy. It follows that the tax subsidy is collected in the amount of incremental receivables (excess invoices). The previous tax subsidy accumulated the previous cash flow of enterprises and has paid the previous volume of invoices; the tax subsidy of the new period pays the incremental volume. The simple matrix below illustrates this point:

Current volume of invoices during payment period	
Previous volume of invoices during payment period	Incremental invoices
Previous cash flow of enterprises during payment period	The tax subsidy
Current cash flow of enterprises during payment period	

This relationship gives the rise to an incredible fiscal system in which

excess invoices = the tax subsidy.

This literally means that if Peter charged Paul an extra $1,000 for a pile of sawdust, between the two of them they can readily collect the said $1,000 from the Russian Uncle Sam (or whatever his proverbial name might be). And this is not the end of lunacy. Recall that the tax subsidy is off-budget. By its very nature, it cannot be financed by tax revenues because tax revenues are what enterprises remit in cash, whereas the tax subsidy is what enterprises do not remit in cash or remit only in exchange for Central Bank credit. The latter represents a quasi-fiscal subsidy and adds to the tax subsidy. The tax subsidy is thus an expenditure in excess of fiscal revenues. It falls into the budget deficit and is financed by government debt, which consists of money and bonds. For reasons which will become evident shortly, the government cannot in most cases run additional budget deficit and debt. Suffice it is to mention that, whenever enterprises sense the government capacity to sell additional debt, they increase the volume of excess invoices, take a bigger tax subsidy, and thus absorb the funds raised through the debt. Enterprise managers who survive, indeed even prosper, have learned to behave in this way. This means that the **government budget net of the tax subsidy is, most of the time, balanced**. This is both because the tax subsidy is financed by debt, not taxes, and because the government can rarely finance other expenditures by debt. This is not the relationship that holds every day or even every month, because the government can from time to time make sudden changes. However, over time it holds firmly. The following most incredible fiscal system emerges:

excess invoices = the tax subsidy = added government debt.

Or, in stock terms,

receivables = accumulated tax subsidy = government debt.

The world has never seen anything like it; the world has actually not been ready for it, and thus missed the warnings of the impending Great Default.[24] Figure 1.1 depicts the relationship between the stock of receivables and Russian government debt in the combined form of internal bonds and money from 1994 through August 1998, when the Great Default happened.[25] We drop the tax subsidy here and take a shortcut comparing directly the initial force—excess invoices—and the final derivative, government debt. Compare them ruble for ruble, so to speak. They match closer than we expected, despite the fact that these are crude approximations with numerous measurement problems.[26]

---

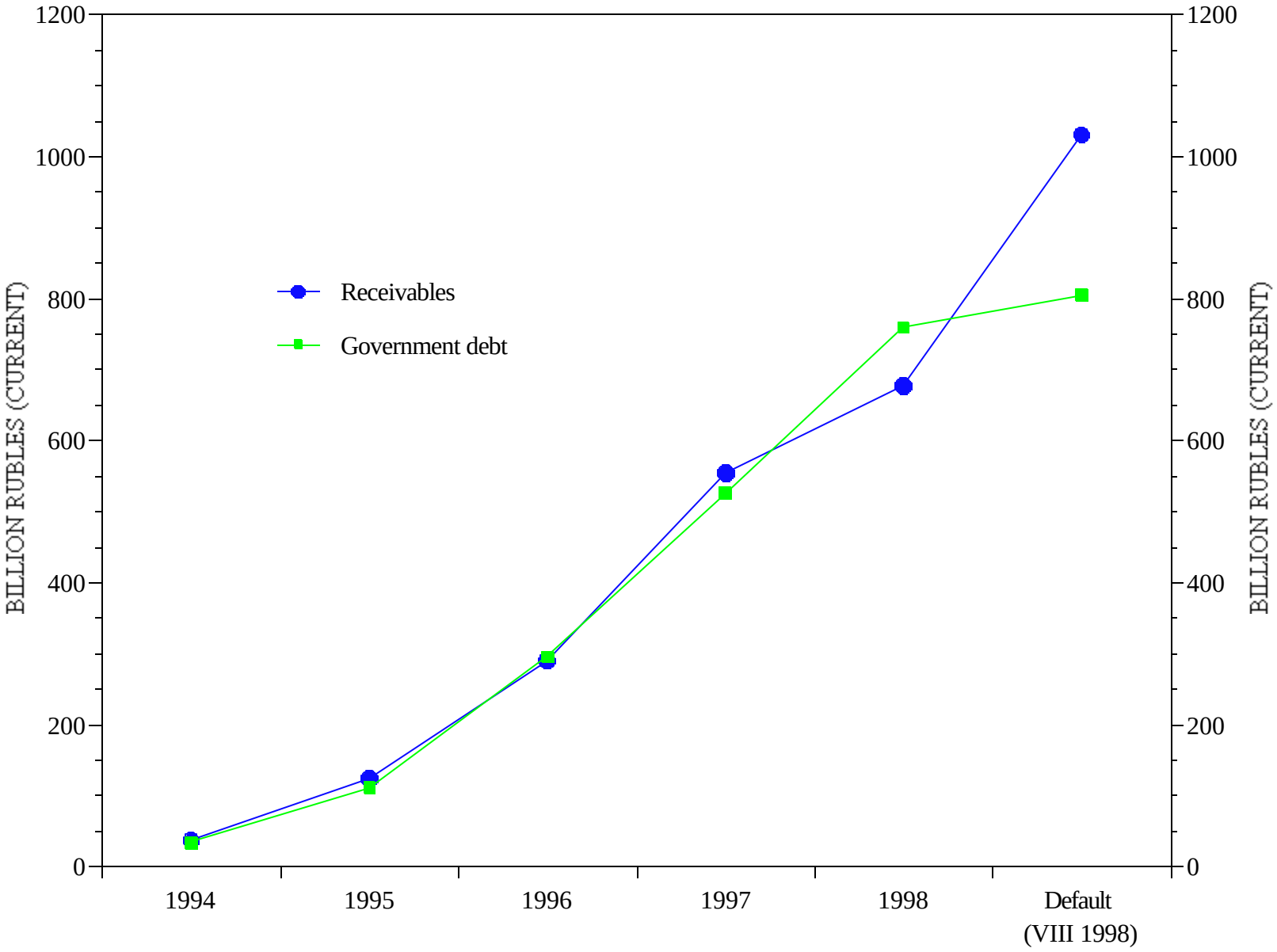
[24]Michael S. Bernstam and Alvin Rabushka, *Fixing Russia's Banks*, pp. 99-108 and "Russia's Banks Need to be Reformed, Not Rescued," *The Wall Street Journal*, July 15, 1998.

[25]We will show in chapters 9 and 10 that the same relationship holds in 1992 and 1993. We omit the earlier years here for a presentational reason: The amounts are presented in nominal rubles. Due to the high inflation in 1992-93, the amounts in those years can hardly be seen on the diagram which covers the period 1992-98 using a linear scale.

[26]Chapters 9, 10, and 11 undertake a thorough accounting decomposition of the tax subsidy and government debt. They present separate and detailed empirical comparisons between the tax subsidy and excess invoices and the tax subsidy and government debt. They employ the flow measures in addition to the stock measures. For the current shortcut presentation in this chapter we use the stock form of presentation. This is both because it smooths short-term fluctuations and because the official data is in stock form; the flows are our calculations with measurement problems of their own. The measurement of the accumulated stock of excess invoices is straightforward. It is the current stock of enterprise receivables. The measurement of government debt raises a panoply of caveats. We include internal bonds only, such as GKO and OFZ, and ostensibly exclude the external debt. In fact, we do include the relevant part of external debt. Part of the external debt constitutes old Soviet debt and accrued interest on its arrears. This debt is not related to financing the tax subsidy. The other part of external debt, such as Russian Eurobonds and some IMF loans, is implicitly included in our combined measure of government debt. This is because the stock of money does contain the currency issued when the Central Bank purchased dollars from the government, the dollars raised through external borrowing. Other parts of external debt, that to the IMF for foreign exchange reserves, is not included in our measure of government debt and is not relevant for studying the impact of excess invoices on the tax subsidy and public debt. Our measure of internal bonds is imperfect on at least two counts. For the lack of data, we do not include very short-term bonds, such as KOs and similar government "junk bonds" and promissory notes with the less than three-month and often a few days maturity. One valid justification for their exclusion is they are often used as tax offsets and thus eventually end up in additional budget deficit and regular bonds. A more serious weakness is that we did not exclude, for the lack of clear data, the bonds purchased and held by the Central Bank. This creates an overlap between two forms of debt, namely bonds and money. The most difficult measurement problem in our diagram is accounting for money as part of internal debt. Measuring money is a perennial problem even for specialists, well beyond the field of our book. Our measurement of money as government debt is atypical. In our measurement, money as the quasi-fiscal subsidy and the quasi-fiscal debt includes not only the amount of money printed by the Central Bank (what specialists call the monetary base) but also the deposits multiplied by the banking system from Central Bank credit to enterprises and monetization of the budget deficit. This addition of deposits indicates a peculiar subsidy multiplier for enterprises. This multiplier is not equal to the familar deposit multiplier of the money stock (the ratio of the monetary aggregate of M2 or

FIGURE 1.1

EXCESS INVOICES AND GOVERNMENT DEBT, RUSSIA, 1994-98



Note: Government debt consists of money and domestic government bonds GKO and OFZ. It does not include other domestic bonds, with less than three-month maturity. The stock of money implicitly includes part of the external debt, which the government sells to Central Bank, for which the latter prints currency.

Sources: Government bonds and enterprise receivables: Russian State Committee on Statistics

Money: Central Bank of Russia

One can observe in figure 1.1 that the growth of excess invoices, congealed in the stock of receivables, accelerated over time in 1994-98, except for the second half of 1997, and exploded in the first half of 1998. The growth of government debt moved along the same trajectory, but slowed down in the first half of 1998. Thus came the time when the government could no longer place additional bonds to cover the growing tax subsidy and its commensurate true budget deficit. There is always an upper bound at which the public is willing to hold government bonds. After this upper bound has been reached, a default occurs in one or another form, usually an implicit default, when the government prints money to monetize the debt.[27]

The government could print money and substitute one form of debt, bonds, with another form, money. That option was not feasible because the Russian government ran a pseudo-fixed (pegged) exchange rate, and printing more money would have crashed the currency even before the devaluation of August 1998 (simultaneously with the Great Default). An early devaluation would have led foreign and domestic bondholders to dump bonds. This, in turn, would have left the government no other option but more monetization. Replacing the bulk of the bond stock with freshly printed money would have led to more than a mere hyperinflation—to a complete loss of currency, the tax base, and the ability to spend, when the population would have shifted to dollars as the currency of choice and abandoned rubles. The real choice was between repudiating government bonds before, or at the same time with, devaluation, and government abdication and chaos.

The Great Default of August 1998 took scholars, investors, and Western leaders by surprise.[28] Investors were watching the fundamentals, such as the budget deficit (a meaningless fake) and inflation (which was declining at the time, like the heat is declining in a mirage).[29] Who would have thought to watch

---

M3 to the monetary base), which depends on the demand for deposits. By the rule of thumb, the subsidy multiplier is about 2, because for each ruble of credit transmitted from the Central Bank, the banks create a ruble in deposits, and the 2 rubles pay for excess invoices. The deposit multiplier is an unrelated matter, and it can be as high as 10 (in Germany and Japan) and 11 (in the U.S.). It so happens that is Russia the demand for deposits has been extremely low, because households hold their savings in dollars under the mattresses, and the deposit multiplier hovered around 2 during 1992-1999. This greatly simplifies our job here and allows us to take the monetary aggregate M2 as a crude approximation of the quasi-fiscal subsidy and the quasi-fiscal debt.

[27]For a brilliant theoretical discussion see Thomas J. Sargent and Neil Wallace, "Some Unpleasant Monetarist Arithmetic," Federal Reserve Bank of Minneapolis, *Quarterly Review* 5, no. 3 (Fall 1981): 1-17.

[28]Even the most sophisticated investment houses, including the one run by Nobel prize winners in economics, and international financial organizations lost huge sums of money in Russia's Great Default.

[29]The post-mortem literature adds such explanations as the global financial crisis and the inherent difficulty of post-Communist transitions. See Thierry D. Buchs, "Financial Crisis in the Russian Federation: Are the Russians Learning to Tango?" *The Economics of Transition* 7, no. 3 (1999): 687-715; Erik Berglof, *Stuck in Transit: Rethinking Russian Economic Reform* (London: Center for Economic Policy Research, 1999); Tuomas Komulainen and Iikka Korhonen, eds., *Russian Crisis and Its Effects* (Helsinki: The Bank of Finland, BOFIT Institute for Economies in Transition, 2000); and, *OECD Economic Surveys 1999-2000, Russian Federation* (Paris: OECD, March 2000), pp. 34-45.

an obscure, yet the most meaningful, statistic: the relationship between excess invoices and government debt? When the former greatly exceeded the latter, the tax subsidy could no longer be financed, and the Russian financial house fell like a house of cards.

7. Self-Enforcement and Self-Regulation on the Edge

The unique feature of the tax subsidy is that it is self-enforceable. The entire new fiscal system, with its equivalence of excess invoices, the tax subsidy, and government debt, is self-enforceable and self-regulating. Let us look in more detail how it works on the edge of halting payments: what the government can and cannot do, and what enterprises can and cannot do.

The government can always try to break the continuity of the tax subsidy. It can always try to enforce more tax remittance. This means, it can try to reduce the tax subsidy in various forms, such as tax non-remittance, tax offsets, tax exemptions, Central Bank credit to enterprises, etc. This moves part of the cash flow from payments between enterprises to tax payments. On the edge of halting payments, this shift automatically jams the traffic of payments across enterprises and threatens the production flow and the tax base. For this reason, the government seldom tries to enforce tax remittance and always retreats when it tries.[30] It would rather default on its debts, as it periodically does. The perennial Western insistence that the Russian government increase tax collection[31] is oblivious of the facts: The taxes have been fully collected from the public but not fully remitted by enterprises; and by forcing more remittance the government will self-destruct. The latter makes the tax subsidy self-enforceable.

The mechanism of self-enforceable tax non-remittance on the edge of a payment jam is automatic. Let us consider various possible situations. We will observe the convergence of their results to the same initial position, to the same payment jam equilibrium:

The singularity of the Great Default in Russia renders these explanations insufficient.

[30]We list in Part Two a number of examples, which amount to unintended experiments, confirming this point. The latest episode occurred during April 3-14, 2000, when the government tried to reduce tax offsets and tax non-remittance on the part of the two largest Russian enterprises, the electric power generator and utility, the Unified Energy System, and the natural gas manufacturer and transporter, Gazprom. It ended up with Gazprom paring down its supplies to the Unified Energy System and the latter turning off electricity supplies to industries and cities. After this episode, the government allowed them to raise controlled domestic supply rates by 21 percent and 35 percent, respectively, thus compensating for the past, present, and future increases in tax remittance. This compensation is to the tune of R15 billion ($530 million) in annual revenues to each enterprise, or 0.6 percent of GDP between the two of them, just as an **additional** subsidy on top of the current subsidy. See Jeanne Whalen, "Russia to Lift Rates to Aid Monopolies in Utility Squeeze," *The Wall Street Journal*, April 14, 2000, p. A16 and Russian Business Consulting, at http://www.rbc.ru/news/free/2000 regular releases during April 3-14,and April 28-30, 2000.

[31]"The thrust of Fund advice to Russia is to strengthen the government's finances by better tax collection." (John Odling-Smee, Director of the International Monetary Fund's European II department, "A Letter to the Editor," *Wall Street Journal Europe*, August 24, 1999). "Mr. [Stanley] Fischer [First Deputy Managing Director] added that Russia's biggest economic problem remains tax collection." ("Headline," *The Financial Times*, September 10, 1999, p. 1).

1. Suppose the government undertakes a partial crackdown to enforce tax remittance. It forces selected enterprises to remit full current tax liabilities or taxes past due. The government succeeds at that. Affected enterprises automatically reduce payments to suppliers in the same amount. The latter automatically reduce their tax remittance in the same amount.[32] Losses are equivalent. The government gains nothing.

2. Suppose the government conducts a large or an across-the-board complete crackdown on tax offenders. Payments between enterprises collapse and a chain reaction of shipment stoppages and supply breakdowns begins. Large suppliers of energy, fuel, and other resources halt supplies to non-paying customers. The tax base narrows quickly. The government may face greater incremental losses of revenues due to output contraction than incremental gains from forced remittance of taxes.

3. Suppose the government starts selective bankruptcies. This is a measure many researchers, policy advisors, and Western investors advocate as a solution for Russian fiscal and other problems. The government can readily achieve this. All requisite bankruptcy laws are on the books. The government can enforce them. Selective bankruptcies reduce payments to suppliers who, in turn, reduce their tax remittance and increase their tax subsidy in the same amount. The government gains nothing in the short run and narrows the tax base for the future.

4. Suppose the government reduces Central Bank credit to enterprises for remitting tax revenues. Enterprises increase tax non-remittance ruble-for-ruble of foregone monetization. They lose the modest multiplier that the banking system creates when it makes credits and opens deposits on the basis of freshly printed money. For this reason, their tax subsidy declines. But the government gains nothing even if enterprises lose part of the subsidy.

5. Suppose the government increases Central Bank credit to enterprises for remitting tax revenues. This increases payment between enterprises and tax remittance to the government. But this does not constitute a special gain for the government because it could just as well issue bonds in the same amount and sell them to the Central Bank or simply arrange direct Central Bank credit to the government.

6. Suppose the government increases tax rates or levies new taxes. This reduces the cash flow of enterprises and their mutual payments, either directly or indirectly, through declining consumer

---

[32] Political connections do not matter in this exercise. Suppose the government targets the less politically connected enterprises. They increase tax remittance and reduce payments to suppliers, getting more inputs without paying. This does not worsen their financial position. Suppose better politically connected enterprises escape the government wrath of tax enforcement. They receive smaller payments from their buyers and increase the self-taken tax subsidy in the same amount. Their financial position does not improve. On the edge of the payment jam, political connections turn out to be fungible and their benefits socialized.

demand. Then enterprises increase tax non-remittance. The government may end up with little or no revenue gains.

7. Suppose the government reduces regular expenditures, outside of the tax subsidy, in order to compensate itself for lost tax remittance and to reduce the budget deficit. This is possible up to a point—until the point is reached when enterprises provide inputs for which they are not paid to parties that have lost government payments. These may be the military, non-profit organizations (schools, hospitals, etc.), and households that use public utilities. Then enterprises collect the tax subsidy from the government in the amount of unpaid supplies.

8. Suppose the government increases regular expenditures and issues more debt to finance the growing budget deficit. Enterprises observe the increased debt-financing by the government. They respond by increasing the amount of unpaid invoices to each other, inflating invoices, and collecting a higher tax subsidy by reducing tax remittance. They crowd out government expenditure whenever the government raises more funds through debt. The combination of the last two paragraphs indicates that the budget, net of the tax subsidy, is balanced most of the time. During the time when enterprises fail to catch up with government moves, the budget is in surplus. Only in exceptional circumstances can it be in deficit; in practice, we did not find such exceptions for any quarter during 1992-1999.

9. In addition to self-enforcement of the tax subsidy, self-regulation is also at force. Self-enforcement limits what the government can do. Self-regulation limits what enterprises can do. Suppose enterprises increase tax non-remittance, and take a higher tax subsidy than they need for payments. In this case, they have more cash flow to increase payments to suppliers and mitigate the payment jam. After that, the government can enforce more tax remittance without jeopardizing production flows and future tax flows. An increase in the current tax subsidy reduces the future tax subsidy by the same amount.

It follows that fiscal policy and monetary policy are powerless under this fiscal system.[33] This conclusion is simply another way of saying that the tax subsidy is self-enforceable and self-regulating in the payment jam, on the edge of halting production and tax flows. If the government deviates or enterprises deviate from the level of tax subsidy needed to meet enterprise payments, they automatically bounce

[33]We do not claim that enterprises were able to take over all the economic powers of government. The Russian government was able to make discretionary grants of assets at below-market prices to favored individuals and enterprises, as in the infamous loans-for-shares scheme we discuss in Part Two. Enterprises fought among themselves for favored treatment on the part of government officials. It may well be that the perilous fiscal situation prompted the government to accept the absurd terms offered in the scheme. But it was the government, not the enterprises, that allocated the specific assets. This case of government discretionary authority is more the exception than the rule under the enterprise network socialism that emerged in Russia in the 1990s.

back.[34]

Self-enforcement and self-regulation of the tax subsidy demonstrate the automatic mechanism through which the tax subsidy equalizes with excess invoices and creates in turn an equal amount of government debt. This mechanism operates the summary equation presented above. When enterprises create excess invoices they increase the tax subsidy by the same amount. The government debt increases by the same amount. Figure 1.1 illustrated this self-enforcing and self-regulating mechanism with actual data. Suddenly, the obscure phenomenon of the payment jam and excess invoices, which does not even have a name, rises as a ghost behind the Great Default.

8. The Accounting Mechanics of the Great Contraction

Excess invoices have accounting consequences for output. Invoices embody nominal output (nominal income), a bundle of real output at some price level. Buyers order this output, sellers produce and ship it, and attach invoices. Cash flow pays for this output during a payment period, in the amount of invoices. The final output is purchased by consumers; invoices ultimately meet the entire cash flow in the economy. What managers and accountants see as cash flow, the specialized literature views as spending—the money stock times its velocity of circulation. The latter, also known as income velocity of money, can be thought of as the number of times the money stock turns over during a calendar period, usually one year, to pay for the nominal output produced during this period.

A greater amount of invoices relative to cash flow takes a longer period to pay off. This disparity, excess invoices, may have vastly different reasons, on the cash flow side or on the invoice side:

- Cash flow may decline—the money stock or its velocity, or both. Banks may crash and deposits dwindle. Or, under a fixed exchange rate, in response to capital outflow, the currency board sells dollars and withdraws local currency from circulation. Or the money stock remains unchanged but consumers slow down spending. This decreases money velocity.

- Enterprises raise prices in invoices in order to enforce the tax subsidy. Enterprises have self-

[34]Economic self-enforcement leaves little room for extraneous political factors. Popular and part of the scholarly literature attribute Russian economic decline to adverse political factors rather than the failure of economic policy. The most widely-touted causes include the lack of political will, the weak state, and the obstructive Parliament. They are summarized most lately in Andrei Shleifer and Daniel Treisman, *Without a Map: Political Tactics and Economic Reform in Russia* (Cambridge, MA: The MIT Press, 2000). But no amount of political will can reduce the amount of a self-enforceable subsidy. The same can be said about the notion of the weak state. This is a weak concept. Like the lack of political will, the weak state does not lend itself to scientific scrutiny. How can we know that the state is weak or strong or that political will is or is not sufficient? Neither category can be measured and tested, verified and falsified. These concepts can explain away any economic or astronomic failure and thus explain nothing. The notion of the Parliament's obstruction of government efforts is a canard. The Parliament is a scapegoat. The paper budget is a fake, the budget net of the tax subsidy is usually balanced, and the budget deficit, public debt, and defaults depend on the tax subsidy collected by enterprises, not on taxes and expenditures legislated by the Parliament.

enforceable fiscal expectations (the tax subsidy). They include the long-term expectations that the government will print money to cover fiscal shortfalls and unsustainable debt. Price increases are automatically built-in in the rising nominal volume of invoices because invoices price shipments. Price increases in excess invoices represent self-fulfilling inflationary expectations. The price increase passes onto consumers, which makes excess invoices relevant in comparison with total cash flow in the economy.

To peruse the traffic jam metaphor, either the road narrowed (the money stock dwindled) or stop-lights slowed the traffic (velocity declined) or more cars entered the road (price increases raised the nominal volume of invoices). Whatever the reason behind excess invoices, the outcome is identical: *The same bundle of real output is purchased and produced during a lengthier period.* As for a calendar period, say, 365 days, it sees a smaller bundle of real output.[35] As in the traffic jam, fewer shipments are delivered every day. Excess invoices thus either incarnate or generate contraction.

For example, suppose the money stock is 400 rubles and its velocity is 2.5 during the year. It can pay for 5 bundles of real output invoiced at 200 rubles each. Suppose the money stock falls to 360 rubles; or the money stock remains unchanged but its velocity declines to 2.25. Only 900 rubles of annual cash flow face 1,000 rubles of invoices carrying 5 bundles of output. Or suppose enterprises raise prices and invoice 222 rubles per bundle. The 5 bundles now require 1,111 rubles of payment while annual cash flow remains 1000 rubles. It will take 405.5 days instead of 365 to pay off invoices to buy the same 5 bundles of output. During the year, only 4.5 bundles of real output will be paid for and produced. This is a 10 percent contraction.[36]

We can write:

cash flow over time/invoices over time = the index of real output

This is a purely mechanical relationship. It tells us nothing about the actual behavior of enterprises, the government, and households: who and how launched this contraction. Moreover, it cannot distinguish between the case of post-Communist economies, in which excess invoices are relevant, and other cases, which may occasionally happen in Western market and developing economies and in which excess invoices are utterly irrelevant. Such are the above-mentioned cases of the banking crash, the currency withdrawal due to capital outflow under a fixed exchange rate, and the decline in velocity due to slower consumer spending. In these three cases, the excess of invoices over cash flow is merely a secondary and uninteresting consequence of the overall contraction, which had nothing to do with invoicing. Excess

[35]The real demand declines. Diagrammatically, the demand curve shifts left, lowering output.

[36]Note how a small increase in the obscure matter of the length of paying-off the average invoice, just 8 days (from 73 to 81 days), which most literature would dismiss as beneath discussion, can produce such a great effect. Six years of such payment lengthening can contract the economy almost by half.

invoices are unimportant in these cases: The duration of excess invoices could last one payment period only or less—a month, a month-and-a-half at most. By that time or even earlier, firms adjust downward their shipping orders and invoices to match the reduced cash flow.

Excess invoices can be relevant by themselves only if they make an *independent* influence on output. This happens if (as in the case of Russia and similar post-Communist economies) they increase when the money stock and velocity do not decline and if their excess continues for quarters and years, not days and weeks. This is an empirical question. The empirical evidence can reveal the behavior of enterprises, which commands invoices; the government, which commands the money supply;[37] and households, which command money velocity.

If the above mechanical relationship holds empirically over time, excess invoices are important. Otherwise, the relationship would not hold longer than one-and-a-half months. It would even be absurd because in Western market, developing, and central plan economies the ratio of the cash flow to invoices stays stable while output grows (and occasionally declines). Excess invoices are impossible, like the payment jam is impossible, outside of the world of the tax subsidy. Under central planning, the government automatically pays residual enterprise bills. It prints money and matches cash flow with invoices at each point in time in order to enforce production, output mix, and designated delivery. In Western market and developing economies, sellers would not ship goods and invoice customers unless payment is expected without delay. Hence cash flow matches invoices at each point in time. Buyers, on their part, would not order inputs unless their expected cash flow matches invoices. The payment period also remains stable, about a month or slightly more. Exceptions mentioned above are extremely rare, short-lived, and unimportant. Only in Russia and similar post-Communist economies can excess invoices exist. They exist because the tax subsidy is collected by enterprises to pay them off, at lengthier and lengthier intervals. Then this relationship can hold over time.

Its nature is the separation of spending from payment, unique to post-Communist economies. The definition of spending as the product of money and its velocity automatically and traditionally treats spending as payment. There can be no disparity between them at any period of time, no excess invoices. This is natural in market, developing, and central plan economies. Looking at invoices in post-Communist economies as spending separated from payment, and looking at cash flow as genuine payment (money times velocity) allows us to see the unusual mechanics of contraction. We can think in terms of two different kinds of spending. One is embodied in invoices and separated from payment. The other is embodied in genuine payment, which is cash flow. By analogy with counterfeit money, one can think about spending, which is separate from payment, as counterfeit spending. Real payment represents the one and only genuine spending. Counterfeit spending orders and invoices goods and builds-in price increases without automatically absorbing price increases, as velocity would do in payment. This is how excess

[37]Banks in Russia play a very small role in creating money by deposit multiplication, and household demand for deposits is extremely low. The deposit multiplier hovered around 2 and rarely approached 3 during 1992-2000. For a detailed discussion, see Michael S. Bernstam and Alvin Rabushka, *Fixing Russia's Banks*.

invoices can mechanically exist.

The unwritten premises of spending as we know it in market, developing, and central plan economies, and in all world economies since time immemorial, since the first trader said to the first producer "I owe you," thus accepting the first unwritten invoice, are these two:

1. One man's invoice is another man's payment.
2. One man's spending is the same man's payment.

These premises have changed in Russia and similar post-central plan countries to these two:

1. One man's invoice is the tax subsidy (or another man's payment after the tax subsidy).
2. One man's spending is the tax subsidy (or this man's payment after the tax subsidy).

Excess invoices, paid off by the tax subsidy, separate counterfeit spending from genuine spending. We can rewrite the above relationship to add substance to the mechanics:

accumulated genuine spending/accumulated counterfeit spending
= cash flow over time/invoices over time = the index of real output

We can now test this relationship with Russian annual data for the period 1992-1999 plus the first quarter of 2000. Figure 1.2 plots the index of annual real GDP versus the ratio of money to receivables at the end of the same year.[38] We disaggregate cash flow into its two constituent components, money stock and velocity, and then relate money stock alone to receivables. As a result of disaggregation, we can observe the behavior of the government and households, which determine money stock and velocity respectively, and the behavior of the enterprise network, which determines excess invoices.

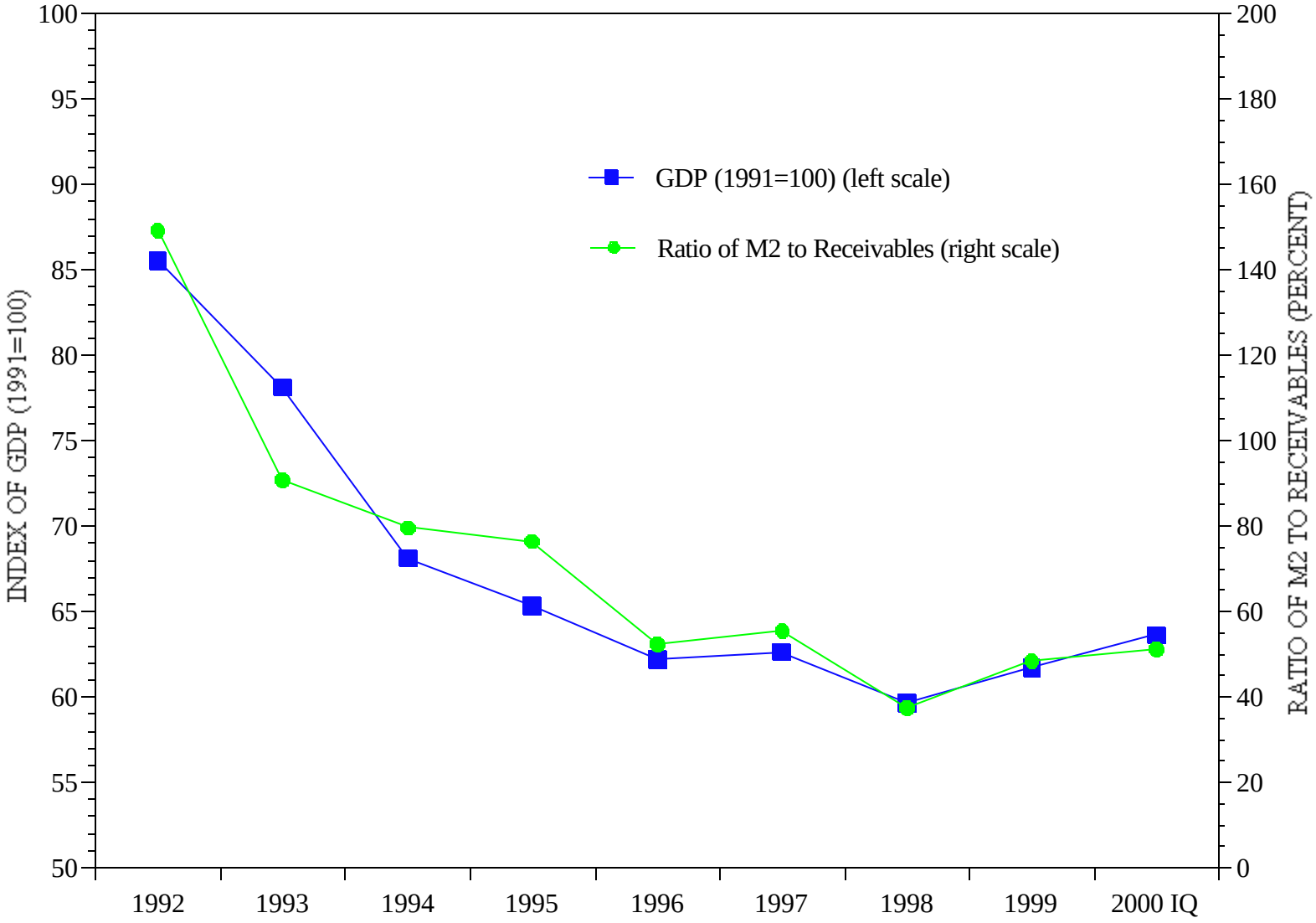
Excess invoices are measured, as in figure 1.1, as the stock of receivables. The money stock is measured as the monetary aggregate M2 (the sum of currency in circulation and ruble deposits).[39] The curves of GDP index and M2 over receivables have different scales. The juxtaposition of the scales

[38]We excluded the comparison of changes between the output index and the ratio of cash flow to receivables from 1991 to 1992 because of numerous measurement problems and administrative write-offs of receivables in 1992. We extend the monthly series to the beginning of 1992 in Chapter 16. However, for completeness, we retained here the output index as of 1991 (that is, GDP in 1991 is taken as 100), to analyze the path of contraction since the abolition of central planning. Also, we do not use any lags for output in relation to cash flow and receivables because the lags varied over time and we want to avoid arbitrary choices and data manipulation.

[39]Arguably, a better measure of the money stock is broad money, the sum of M2 and dollar-denominated deposits, which, in Russia, are chiefly enterprise, not household, deposits; dollar deposits represent additional demand deposits used in transactions. Unfortunately, the data on broad money is available only since mid-1995. However, additional diagrams, which we present in Chapter 16, use broad money along with M2, in relation to receivables, and the results are of a similar nature (but of different values).

FIGURE 1.2

INDEX OF GROSS DOMESTIC PRODUCT (GDP) (1991=100) AND THE RATIO OF M2 TO RECEIVABLES, RUSSIA, 1992-2000



Note: The difference between the scales of the two axises indicates the change in the speed of consumer spending (money velocity).

Sources:

Gross Domestic Product and enterprise receivables: Russian State Committee on Statistics;

The monetary aggregate M2: Central Bank of Russia

accounts for the increase in velocity. This has an empirical reason. The two scales, taken together, show over time a velocity increase by a factor of 4, from about 1.5 in early 1992 to fluctuating around 6 since 1995 through the first quarter of 2000 (divide 6 by 1.5).[40] We have chosen the specific scales on the left and right axes to explicitly incorporate this fourfold increase in velocity between 1992 and 2000. This scalar factor of 4 shows that a major velocity increase, typical for periods of big inflation (when households reduce their real money balances to minimize the loss of their value), partly compensated for a sharp decline in the ratio of money to invoices, and thus cushioned the output decline. Major velocity increases occurred in 1992-93, at the peak of inflation. Velocity increase continued through 1994, after which velocity relatively stabilized and hovered around 6. The two curves have similar shapes. In figure 1.2, they accidently coincide with each other. This is not necessary at all. The relationship is shown not by their coincidence, but by the correspondence of their change year-to-year and over the entire period after taking the change in velocity into account.

In general, figure 1.2 shows a consistently close relationship between the ratio of money to invoices and the index of real output during 1992-2000. Their close association in both upward and downward annual fluctuations in 1997, 1998, and 1999 demonstrates that this is not a mere coincidence of two secular trends, going down for their own separate reasons. These fluctuations, especially upward movements in the ratio of money to receivables and short-term output recoveries, reveal the mechanics of the relationship. Excess invoices run the system. When enterprises have income windfalls, primarily due to an increase of world prices of natural resources, as in 1997 and 1999-2000, they temporarily slow down, stop, or even reverse excess invoicing. They cannot create a payment jam on the edge of halting the payment, production, and tax flows when they have extra income and cash flow, and the government can enforce more tax remittance. During these spells, the money stock exceeds the stock of receivables, payment periods shorten, and we can observe the spikes of output recovery in figure 1.2.

Over the entire period, the ratio of cash flow to invoices and the trend of output must continuously decline. The government prints money as part of paying off the tax subsidy and households initially increase money velocity. Both freshly printed money and rising velocity absorb a commensurate part of price increases built-in in excess invoices and mitigate contraction. But the nature and self-enforcement structure of the tax subsidy are such that money always constitutes only part of the tax subsidy. Tax non-remittance is another part (there are also payroll arrears and other items). The government minimizes monetary accommodation and enterprises use tax non-remittance to enforce it. Only printing money adds to the cash flow; tax non-remittance merely redistributes income and cash flow, smoothing enterprise payments. As the enterprise network continuously creates excess invoices, cash flow lags behind. Output continues to contract.

If the government prints more money, enterprises sense a higher tax subsidy opportunity and raise excess invoices faster. Enterprises are free to charge. No amount of monetization can catch up with excess

[40]If velocity did not change, the tested relationship of payment to receivables would consist of just money and receivables. There would be no need to derive a scalar factor to account for the change in velocity.

invoices except in a very short term. Hence, the ratio of money to receivables (and the ratio of cash flow to receivables, even when money velocity increases) generally and continuously declines, except for short-term reversals due to external income windfalls. The implication is a protracted contraction, not a short-term recession, not a one-time depression, but a Great Contraction. If the relationship holds, this contraction can be endless.[41]

9. The Total Subsidy

The tax subsidy redistributes income from the public at large to enterprise owners and managers. But before this happens at the end of the day, another redistribution takes place during business hours. The very process of the government and the public paying off enterprise invoices begets counterfeit spending, which redistributes income between enterprises. Counterfeit spending, excess invoices, and their subsidized payment permeate the transactions between enterprises and industries across the entire economy. Some of the sellers receive more income than market forces would warrant. Buyers obtain part of their inputs for free. This arithmetically means that some of the sellers are paid less for the resources than market forces warrant. Counterfeit spending and public payment for it create the subsidy across enterprises and industries. This is an invisible web of cross-subsidies and cross-taxes, which necessarily accompanies the tax subsidy.

As in every redistribution, there are winners and losers. Among individuals or households, the winners in Russia and similar post-Communist economies are enterprise owners and managers (including government officials as managers of the process). The losers are the rest of the public.[42] Among enterprises and industries, the losers must be producers of natural resources and high value-added output: There is simply no one else there who can subsidize other producers. Note, of course, that these are impersonal losers, or losers in the purely accounting sense. Their owners and managers are still winners and may even be (and, in the case of natural resource enterprises, usually are) the biggest winners.

These simple and non-controversial accounting observations carry rather contentious implications, which may indeed be anathema from the standard viewpoint:

---

[41]The excellent literature on arrears made one telling error. It assumed that the government would be forced to print money in the full amount of overdue invoices. The literature overlooked the tax subsidy and its composition of money and tax non-remittance. The predicted outcome of complete monetary accommodation of arrears is a combination of high inflation and output stagnation, a low-level equilibrium trap, not protracted contraction with variable inflation, a perpetual downfall. See Guillermo A. Calvo and Fabrizio Coricelli, "Credit Market Imperfections and Output Response in Previously Centrally Planned Economies" and "Inter-Enterprise Arrears in Economies in Transition"; Enrico C. Perotti, "A Taxonomy of Post-Socialist Financial Systems" and "Inertial Credit and Opportunistic Arrears in Transition"; and, Fabrizio Coricelli, *Macroeconomic Policies and the Development of Markets in Transition Economies*, pp. 52-85.

[42]It is an empirical question if some workers are net winners if their industries and enterprises gain from redistribution and the tax subsidy trickles down to these workers, even though they at the same time bear the burden of the tax subsidy as taxpayers and consumers.

1. Enterprise income depends on the distribution of the tax subsidy and cross-subsidies, not on market forces of supply and demand. Prices are largely free of government control, but, contrary to the conventional view, free prices do not necessarily constitute market prices. Free prices can embody cross-taxes and cross-subsidies, as they do in the post-Communist Russia. They are still fiscal prices, like they were under central planning when prices were set by the government. Less government and more liberty do not necessarily mean more market, or market at all. A liberalized and privatized economy is not necessarily a market economy. It can still be a socialist economy in the sense of the ubiquitous redistribution across-the-board. Freedom from the government does not necessarily create a free market. It can produce free-for-all socialism.

2. The empirical extent of this socialism may vary. In Russia in 1999, the stock of receivables, which entailed cross-subsidies, was equal to 40 percent of GDP and some 27 percent of total sales, with the average length of payment about 3.5 months and the velocity (turnover) of 3.4 payments per year. This implies that 90 percent of income flow in the economy was redistributed, or socialized, and even more if one considers the transmission of redistribution through consumer prices. The extent of socialism is about the same as it was under central planning, when a small share of income flow was private and largely free of redistribution. As a result of liberalization, privatization, and other reforms, one socialism replaced another. A socialist mutation took place.

3. Profits, except for some profits from exports, depend on the distribution of the tax subsidy and not on the market. Prices embody cross-taxes and cross-subsidies, socialize income, and do not send market demand signals. Therefore, productive incentives and mechanisms are lacking. Economic growth in this system, beyond occasional mechanical and small-scale recovery, is impossible. In this sense, post-Communist economies in Russia and similar countries are inherently less productive than central planning. At least under central planning, production was forced by the government and a medium-term economic growth could be, and had been, achieved.

4. Natural resource enterprises subsidize most other industries. Users of resources can obtain more inputs than they would under market conditions on two counts: The tax subsidy pays for counterfeit spending in excess of enterprise payment from their own earnings, and prices embody cross-subsidies. With these subsidies and free inputs available, many manufacturing, construction, and agricultural enterprises can afford to produce output whose market value in world prices would be lower than the market value of resource inputs. This means that value is subtracted across many industries in a significant part of the economy. The negative value-added, the most notorious feature of central planning,[43] is still being produced. The economy may be as wasteful after the

[43]For a definitive treatment of negative value-added see Ronald I. McKinnon, *The Order of Economic Liberalization. Financial Control in the Transition to a Market Economy* (Baltimore and London: The Johns Hopkins University Press, 1993), pp. 162-186.

abolition of central planning as before, in the sense of subtracting value.[44]

5. The cross-subsidy, which subsidizes value subtraction and other inefficient production, is hidden in invoices and subsidized payment. As a process, it precedes the tax subsidy. As a magnitude, it significantly exceeds the tax subsidy. It redistributes the entire GDP and beyond. Like the tax subsidy, it is off-budget, invisible to budget-watching eyes. It is borne by consumers through cross-taxed prices of final output. The total subsidy combines and, to some extent, overlaps the tax subsidy, which did not exist under central planning, and the cross-subsidy, which did. The scope of the total subsidy in relation to GDP is thus unprecedented by any historical standards.

Ultimately, the extent of redistribution, which defines socialism, is higher in post-Communist Russia than it was in the Soviet Union under central planning. A simple empirical proof is the combination of negative value-added, which embodies the cross-subsidy, and the Great Default, which epitomizes the tax subsidy.

10. Growth Was a Windfall, Contraction an Accounting Impossibility

Value subtraction under central planning carried an invisible growth advantage. Elimination of value subtraction is in itself value addition, that is, one-time economic growth. Other vast inefficiencies of central planning contained inborn opportunities for efficiency improvements and thus for additional growth. Market prices and incentives automatically eliminate value subtraction and other inefficiencies and should—indeed cannot fail to—generate instant growth. This made the lack of substantial economic growth in Russia and elsewhere, let alone the Great Contraction, impossible on accounting grounds.

Arithmetically, subtraction of subtraction is addition. This means that simply closing down the value-subtracting enterprises and industries and reallocating (initially, simply selling on the world market) resources wasted by them, automatically generates one-time economic growth. Its potential extent was substantial. For example, if the negative value-added constituted 33 percent of the value of resource inputs, its elimination could produce an instant 50 percent growth of real GDP (in constant prices).[45]

---

[44]We discuss this issue in detail and present empirical evidence in Chapter 14.

[45]A detailed study of input-output tables for Czechoslovakia in 1986 and Hungary in 1987 by Gordon Hughes and Paul Hare found, after adjusting for output quality, that value subtraction amounted to 34.8 percent and 34.6 percent, respectively. Poland exhibited a similar extent of value subtraction. See Gordon Hughes and Paul Hare, "Competitiveness and Industrial Restructuring in Czechoslovakia, Hungary, and Poland," *European Economy, Special Edition*, no. 2 (1991): 83-110. See also Gordon Hughes and Paul Hare, "Industrial Policy and Restructuring in Eastern Europe," *Oxford Review of Economic Policy* 8, no. 1 (1992), pp. 82-104. Value subtraction of around 25 percent can be discerned from the data on the former East Germany, without accounting for output quality. See "Micro and Macroeconomic Adjustment Processes in East Germany," Deutsches Institut fur Wirtschaftsforschung, *Economic Bulletin* 28, no. 4 (Berlin, June 1991), no. 6 (August 1991), no. 10 (December 1991), 29, no. 2 (April 1992), no. 5 (July 1992), no. 9 (November 1992), 30, no. 2 (April 1993), no. 4 (June 1993). If one adjusts for East German data for product quality, the extent of value subtraction should increase and converge with levels in Poland, Hungary, and Czechoslovakia, that

From this perspective, even 28 percent growth achieved in Poland in the 1990s, after a big contraction in 1990-92 and a subsequent recovery, can be viewed as a success only relative to Russia and other depressed economies. Relative to the inherent growth potential, even Poland, let alone every other economy northwest of China, was a failure. The conventional explanation of China's success is the advantage of backwardness, the catch-up—the adoption of accumulated Western technological knowledge at no cost. An accounting perspective suggests that an even greater advantage is the advantage of wrongheaded industrialization—a one-time jump through eliminating value subtraction. The catch-up requires an effort of investment, training, and application of adopted technology. Even the fastest catch-up takes time. The advantage of wrongheaded industrialization provides effortless growth, with no additional investment and training, at no time—an instant windfall. The more industrially developed a country was under central planning, the more one-time growth windfall it could achieve at no cost. Russia's initial conditions were among the most advantageous on this account.[46]

Elimination of value subtraction and automatic, instant growth were easy not only from an accounting but also from a socio-political perspective. As a matter of fact and accounting, the total subsidy pays not only for value subtraction but also for 100 percent of wages of workers engaged in value subtraction. It is a matter of fact because waged workers, not robots, are working in the value-subtracting enterprises. It is a matter of accounting because value subtracting output is somehow produced. This means that the public pays for the difference between input and output prices (value subtraction per se) and also for wages and profits of producers. The latter can be properly called negative wages and negative profits. These wages and profits are positive for their recipients but negative for the economy as a whole. Negative wages and negative profits are subsidized on top of value subtraction. Therefore, if the market closes down all value subtracting enterprises, the government can tax the public and pay 100 percent of wages to displaced workers for not working and for retraining, and the total subsidy will still be lower than before because value subtraction will not be subsidized. Thus substantial, instant, one-time economic growth can be achieved without making workers financially worse-off, at about zero social cost.

Not only was the Great Contraction not a necessary part of transformation, the opposite is true: The Great Contraction was structurally impossible as a matter of accounting. Economic expansion was, and still is, an inherent part of introducing a market economy after Communism or after the new, post-Communist socialism. It is a windfall. Growth was and still is preordained under a proper economic policy. That it did not happen, and the Great Contraction reigned instead, was a matter of policy choice, the choice of economic ideas, and ultimately, the choice of a paradigm.

---

is, around 33 percent of the value of inputs. A similar magnitude of value subtraction can be deduced from unpublished Russian input-output tables for 1991, on the eve of price decontrol.

[46]Paradoxically, from this perspective, even the end of Russia's subsidization of Eastern Europe and the former Soviet Republics with underpriced energy must have helped economic growth in both Russia and its former beneficiaries. Input pricing at world levels should have eliminated value-subtracting, not value-adding, output, and thus contributed to growth, not contraction.