W.J. Usery Workplace Research Group Paper Series

Working Paper 2009-6-2
June 2009

Does Labor Supply Respond to a Flat Tax? Evidence from the Russian Tax Reform

Denvil Duncan
Georgia State University

Klara Sabirianova Peter
Georgia State University

This paper can be downloaded at: http://aysps.gsu.edu/usery/Papers.html

Georgia State University | Andrew Young School of Policy Studies

Does Labor Supply Respond to a Flat Tax? Evidence from the Russian Tax Reform

Denvil Duncan*
Georgia State University

Klara Sabirianova Peter**
Georgia State University
CEPR, IZA Bonn

June 2009

Abstract
We exploit the exogenous change in marginal tax rates created by the Russian flat tax reform of 2001 to identify the effect of taxes on labor supply of males and females. We apply the weighted difference-in-difference regression approach and instrumental variables to the labor supply function estimated on individual panel data. The mean regression results indicate that the tax reform led to a statistically significant increase in male hours of work but had no effect on that of females. However, we find a positive response to tax changes at both tails of the female hour distribution. We also find that the reform increased the probability of finding a job among both males and females. Despite significant variation in individual responses, the aggregate labor supply elasticities are trivial and suggest that reform-induced changes in labor supply were an unlikely explanation for the amplified personal income tax revenues that followed the reform.

Keywords: labor supply, personal income tax, flat tax, labor supply elasticity, difference-in-difference, regression discontinuity, wage endogeneity, employment participation, Russia, transition.

JEL classification: H3, J2, J3, P2

Acknowledgement: We thank Yuriy Gorodnichenko, Barry Hirsch and participants at the Society of Labor Economists Meetings in Boston for useful comments.

* ecodrdx@langate.gsu.edu, Department of Economics, Andrew Young School of Policy Studies, Georgia State University, PO Box 3992, Atlanta, GA 30302-3992, phone (404)-413-0239, fax (404) 413-0145.
** Corresponding author, kpeter@gsu.edu, Department of Economics, Andrew Young School of Policy Studies, Georgia State University, PO Box 3992, Atlanta, GA 30302-3992, phone (404)-413-0165, fax (404) 413-0145.

1. Introduction

Although the classical flat tax - as described by Hall and Rabushka (2007) - has not been implemented in any country to date, a number of countries have chosen recently to tax individual income at flat rates.[1] As of January 1, 2009, the flat personal income tax (PIT) rate schedule was implemented by 24 countries, 20 of which are formerly centrally-planned economies of Central and Eastern Europe and Eurasia.[2] A flat tax reform of this nature is usually motivated by a number of reasons. For example, a flat rate tax simplifies income taxation considerably. Such simplification along with lower rates is expected to lower compliance cost for both the taxpayer as well as tax administrators (Slemrod 1992). Lower compliance cost and lower tax rates could potentially reduce evasion, income shifting, and other avoidance activities.[3] A flat tax reform is also motivated by the desire to stimulate the economy via higher work effort and investment (Hall and Rabushka 2007).

The most popular of these flat tax reforms is the flat PIT rate that came into effect on January 1st 2001 in Russia. Much of the attention received by the Russian reform was due to the dramatic increase in Russia's PIT revenue following the reform. According to Ivanova, Keen, and Klemm (hereafter IKK 2005), real PIT revenue increased by 26 percent in 2001 and continued to grow in subsequent years. From a theoretical perspective, lowering tax rates will affect PIT revenues via two channels. The first is a direct channel resulting from the lower rates,

[1] Unlike the classical flat tax, countries often apply different rates to corporate income, dividends, capital gains, and other sources of income. The common presence of deductions and exemptions also means that the tax is not proportional.

[2] These countries are Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Czech Republic, Estonia, Georgia, Iraq, Jamaica, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Mauritius, Mongolia, Montenegro, Macedonia, Poland, Romania, Russian Federation, Slovakia, Serbia, Trinidad and Tobago, and Ukraine. Several more countries are seriously considering the implementation of the flat PIT rate, e.g., Greece, Hungary, Moldova, and Panama, among others.

[3] Both empirical and theoretical evidence on the link between tax rates and evasion is mixed. However, a majority of the empirical literature find a positive relationship between tax rates and evasion with recent theoretical support for these results provided by Dhami and al-Nowaihi (2007).

which in this case would mean lower tax revenues. The second is the indirect effect on the various components of the tax base due to behavioral responses of individuals to lower rates and other incentives implied by the tax reform.[4] If the second effect dominates, then the tax reform could, in principle, expand the tax base and produce strong revenue performance.

The conclusion that the flat rate expanded the tax base sufficiently to explain the revenue growth has been greeted with much skepticism (Gaddy and Gale 2005 and IKK 2005). IKK (2005), for example, find no evidence that the reform is responsible for the increase in PIT revenue. They also argue that while compliance did increase as a result of the reform, it is not clear whether the lower PIT rate or stricter enforcement should be given the credit. A more definite conclusion is reached by Gorodnichenko, Martinez-Vazquez, and Sabirianova Peter (hereafter GMP 2009) who show that a significant proportion of the increased compliance is due to the lower PIT rate.[5] Besides PIT revenue collections and compliance, the Russian reform has been used to estimate the effect of tax changes on gross declared income (IKK find a negative effect) and net of tax reported earnings (Martinez-Vazquez *et al* (2006) find a positive effect). Both IKK (2005) and GMP (2009) also examine the real side effects of the reform and find little or no response.[6] While the existing evidence has been solid on the compliance and reporting effects of the reform, we believe there is more to be done on the real side responses.

The purpose of this study is to examine, in a careful and methodological manner, the effect of the Russian tax reform on several dimensions of labor supply, including hours of work,

---

[4] It has been shown that there are multiple dimensions along which individuals can respond to changes in the tax rates. They fall into two broad categories: (1) the productivity responses that change the amount of resources available in the economy, including the responses of working hours, employment participation, efforts, job reallocation, etc.; (2) the responses that redistribute the existing resources between various accounts via evasion, income shifting, and other avoidance activities (Slemrod 1992, Feldstein 1995, GMP 2009).

[5] GMP (2009) find that approximately 40 percent of increased reported income relative to consumption can be explained by improved voluntary compliance in response to tax changes, while the remaining part is attributed to other reasons, including better tax enforcement, credit market development, macroeconomic improvements, etc.

[6] IKK (2005) do not find any positive effect of the tax reform on hours of work. GMP (2009) estimate the real productivity effect at the household level using consumption data and find a very small response (0-4% increase).

employment participation, and multiple job holding. Given that IKK (2005) find negative or zero effect on hours and GMP (2009) find little productivity effect, it is natural to ask why look at this question again. We believe there are several reasons for taking another look at the labor supply response. First, the small aggregate productivity effect at the household level can veil significant labor supply effects at the individual level. The consumption-based estimates combine all productivity responses that may or may not have changed in the same direction. The fact that the estimates are based on household level data also implies that the total productivity effect may be hiding opposing effects on individual members of a given household.[7]

Second, the aggregate estimates of the effect of lower tax rates on hours of work as in IKK (2005) may conceal considerable variation in how hours responded by gender, job type, income group, and other individual characteristics. For example, we find significant differences in labor supply response between males and females at the mean and other points of the hour distribution. We also think that it is important to look at the extensive margins of labor supply response that have not been examined in previous studies of the Russian flat tax reform. We find evidence that individuals responded to tax changes along extensive margins such as job finding and multiple job holding.

Third, IKK (2005) focus on the short-term effects of the flat tax reform, as they use only two years of pre- and post-reform data (2000 and 2001). With more longitudinal data being collected, it is important to look at the medium- to long-term effects of the reform on labor supply. We use panel data for two pre-reform years (1998 and 2000) and five post-reform years (2001-2005) and find the response being larger over the longer time span.

[7] For example, one member of the household may have increased hours of work and income while the other worked less (or work the same hours as before the reform). If these effects are relatively the same so that they cancel each other, then GMP's conclusion that productivity effects are small, while accurate on aggregate, would be hiding significant variation.

Finally, there is significant room for improvement of estimates on the methodological front. Although IKK (2005) is the first and only previous study to estimate the treatment effect of the Russian tax reform on hours of work, their estimates represent a simple mean difference in hours between the treatment and control groups. We re-examine their estimates by using the fully specified labor supply function,[8] addressing the issues of tax and wage endogeneity, providing more robustness checks, and applying some of the methodological innovations proposed by GMP (2009) in defining the treatment group and using the regression discontinuity (RD)-type approach.

We identify the labor supply effect using the flat tax reform as a natural experiment. The tax reform eliminated the two top PIT rates replacing them with the bottom rate. This implies that individuals in the top two brackets saw their marginal tax rate fall in the post reform period. We define this change in the tax rate as treatment and estimate the impact of the reform on the labor supply of the treated using difference in difference (DID) and RD-type techniques.[9] In our identification process, we take advantage of the flat tax rate. The fact that all taxpayers face the same marginal tax rate after the reform regardless of income and irrespective of hours of work allows us to escape much of the tax endogeneity issues present in previous labor supply studies.

Estimation is done using micro panel data from the Russia Longitudinal Monitoring Survey (RLMS). Two alternative strategies are used to deal with the endogeneity of wages. The first approach estimates a reduced form labor supply equation derived from a simultaneous system of labor supply and demand while the second relies on labor demand shifters as instruments for wages. Our estimates show that aggregate results hide important variation across

[8] Our estimated labor supply function includes wages, non-labor income, individual and job characteristics, demand shifters, individual fixed effects, and clustered standard errors.

[9] Previous studies of labor supply in other countries applied the DID method in environments that still maintain graduated tax rates (e.g., Eissa and Liebman, 1996). We know of no study that used the RD-type approach in assessing the effect of tax changes on labor supply.

individual groups, especially by gender. Specifically, we find that the reform had a positive and statistically significant effect on male labor supply; males increased their monthly labor hours - actual and usual - by approximately 4–6% following the reform with an implied elasticity of approximately 0.1–0.15. Though quite small, these estimates are within the ball park of male labor supply elasticities (Meghir and Phillips 2008). At the same time, we do not find that females' average hours of work - actual or usual - responded to the tax reform. However, the zero response does not stay constant over the hour distribution. In fact, there is evidence of a positive response to lower tax rates at the bottom and very top of the female distribution of hours. Despite significant variation in individual responses, the aggregate response confirms that the labor supply component of the tax base did not contribute to the increased PIT revenue.

Apart from allowing us to quantify the role played by the labor supply component of the tax base in explaining the increase in PIT revenue, the estimated labor supply elasticities with respect to tax rates are useful for other purposes. Knowledge of the labor supply response is critical for policies such as means tested welfare benefits that depend on how much labor supply responds to incentives. Labor supply estimates also play an important role in general equilibrium macroeconomic analyses that seek to understand how policies – not necessarily taxes – affect economic aggregates. Additionally, most labor supply results exist for developed countries with mature labor markets. Since these results may not apply to transition countries whose labor markets are yet to reach full maturity, direct estimates of the labor supply response - for the largest transition economy - should be informative to the economics profession. A better understanding of the reform's impact on the labor supply component of the tax base should also prove useful to other countries trying to emulate the Russian tax reform.

The remainder of the paper is organized as follows. Section 2 discusses the theoretical implications of the Russian tax reform. Our identification strategy is described in section 3 followed by a discussion of the data in Section 4. We present the results in section 5 and draw some conclusions in section 6.

2. Conceptual Framework

In this section, we briefly describe the Russian tax structure and discuss the effects the reform likely had on labor supply using a static labor supply model. Like many other countries, Russia had a graduated personal income tax schedule before 2001. The tax schedule was progressive with both the marginal and average tax rates increasing with declared income. In one of the boldest tax reforms of the 21st century, the graduated schedule was replaced by a flat PIT rate schedule in 2001.[10] The significance of the reform was later amplified by the dramatic increase in PIT revenue that followed. This made the Russian flat PIT popular among other countries seeking to boost their tax revenue.

The important research question we are interested in is to what extent the increased tax revenue followed the reform is attributed to changes in the labor supply dimension of the tax base. The reform represents a reduction in marginal tax rates which fits neatly into classic models of labor supply with taxes. Since there is nothing novel about the theoretical model, we leave out algebraic expressions and present a simple discussion of the incentives created by the reform. Comparative static results find that the marginal tax rate has an ambiguous effect on

[10] It should be noted that the new tax schedule is not a pure flat tax as outlined in Hall and Rabushka (2007). For example, the new tax structure preserves deductions and applies different rates to dividends and interest income. IKK (2005) discuss in greater detail the different changes that were implemented during the reform. One of those changes that we consider in Section 5 is the unified social tax paid by employers. Specifically, the reform unified all employer contributions to the social security funds (pensions, social insurance, and medical insurance) into a social tax and cut the existing flat rate from 38.5 percent, when all contributions were added together, to a regressive scale with 35.6 percent for income below 100,000 rubles per worker and lower marginal tax rates for higher income levels.

hours worked (Conway 1997). The ambiguity arises from opposing income and substitution effects. A decrease in taxes represents both an increase in income – which causes people to buy more of both consumption and leisure[11] – and an increase in the price of leisure, which causes them to work longer hours. Since the income and substitution effects move hours of work in opposite directions, the net effect depends on which one dominates.

Despite this ambiguity, there are cases where the net effect can be discerned from careful inspection of the budget constraint (Eissa and Liebman 1996). We turn to such an exercise to predict, a priori, the possible effect of the tax reform on hours of work. Prior to the tax reform, the Russian PIT structure had graduated rates starting at 12% for the lowest income bracket up to 50,000 rubles per year and a top rate of 30%.[12] Working individuals also had to pay a 1% contribution to the pension fund. The tax reform eliminated the 1% pension tax and replaced the graduated rate structure with a linear tax schedule that had standard deduction of 4800 rubles and a flat rate of 13%. Figure 1 illustrates the tax schedule before and after reform. While we cannot predict the net effect on hours, the nature of the change allows us to make some predictions for different groups of individuals based on their location on the budget constraint. Consider Figure 2; the nonlinear budget constraint given by ABCDE represents the pre-reform status quo. The tax rates on the pre-reform budget constraint are 1 for segment AB, and 13, 21 and 31 for segments BC, CD, and DE, respectively.[13] The tax reform increased standard deductions from 3168 to 4800 rubles and replaced the two top rates with the 13% rate for segment BC, yielding the post reform budget constraint given by ABF.[14]

---

[11] This assumes that leisure is a normal good.
[12] In 2000-2001, 50,000 rubles were equivalent to $1786 (at the exchange rate of 28 rubles per U.S. dollar).
[13] We adjust the budget constraint for the 1% payroll tax paid by employees.
[14] The higher standard deduction means that the tax rate fell for some tax payers. However, the number of individuals falling into this category is inconsequential. As such, the budget constraint does not reflect the increased standard deduction. Additionally, deductions are not available for all taxpayers. Individuals cannot adjust their income for the standard deduction (400 rubles per month) after their accumulated income reaches 20,000 rubles.

It is clear from Figure 2 that individuals whose optimal choices are on segment AB or BC (excluding the kink point at C) should not be affected by the tax reform. As such, we would expect to see no tax-induced response among individuals in this group. On the other hand, individuals located at the kink point (C) may or may not increase their labor supply depending on their preferences. The standard ambiguous effect from the theory also applies to individuals who are on segments CD and DE. It is not immediately clear what these individuals will choose to do given the change in tax policy. Existing empirical results for this margin are mixed. While the international evidence reveals mostly positive labor supply effects, there is a lot of variation in the size of the estimates (Blundell and MaCurdy 1999, Meghir and Phillips 2008). IKK (2005) find either negative or no effect of the Russian reform on hours of work depending on the definition of the treatment group.

The reform may also have had an impact on participation rates given the lower rate and higher standard deduction. However, three facts lower our expectation of any such effects. First, as mentioned above, the higher standard deduction affected very few individuals. Second, the bottom rate (13%) did not change for the majority of the pre-reform lower bracket individuals. Therefore, any action along this margin would have to be for individuals entering the market at high income levels. Third, in contrast to Central and Eastern European countries, aggregate employment in Russia does not move as much as real wages in response to output fluctuations (Boeri and Terrell 2002 and Gorodnichenko, Peter, and Stolyarov 2009), suggesting that employment participation may not be very responsive to wages.

The implicit assumption that individuals can freely change their hours is not innocuous as employers generally set a fixed package of hours and wages (Altonji and Paxson 1992). Instead of choosing the number of hours to supply at the primary job of employment, workers may

choose to have multiple jobs or moonlight (Matveenko and Saveliev, 2005). Yet, the effect of flat tax reform on multiple job holding is hard to predict. There is a theoretical possibility that if individuals do not pay taxes on their secondary earnings, then lower tax rates on primary wages may induce them to shift hours to their primary place of employment, which implies a *negative* effect on secondary employment (Lemieux, Fortin, and Frechette 1994). However, the flat tax reform may also have a *positive* effect on both the prevalence and reporting of multiple jobs, especially in the Russian case. Prior to the reform, personal income taxes were subject to automatic withholding by employers, and workers with a single job were not obliged to fill out tax declarations at the end of the year. Multiple job holders, on the other hand, were required to submit tax declarations and faced a higher marginal tax rate if their combined tax liability exceeded the amount of taxes withheld by employers. Thus, under the old tax code, individuals would have a lower incentive to have additional formal jobs and a higher incentive to hide (not report) other jobs.[15] After the flat PIT rate is adopted, multiple job holders whose taxes are withheld by employers are no longer required to submit tax declarations, and thus individuals may be more inclined to hold multiple jobs as well as report previously concealed secondary jobs in the survey. Only careful empirical analysis can establish which of these two opposite effects dominates.

In sum, from the theoretical standpoint, the effect of the Russian flat tax reform on labor supply remains ambiguous. Determining this effect becomes an empirical question and faces serious identification problems that will be tackled in the next section.

[15] Because tax is withheld at source, failure to declare a secondary job does not imply that the job is informal. Individuals may hide the secondary job simply to prevent their tax liability from going above the amount withheld. The Russian tax administration at that time did not have capacity to trace individuals with multiple jobs.

3. Identification

3.1. Exploiting the Flat Tax: Difference-in-Difference Approach

In this section, we discuss the empirical approach taken to estimate the effect of the flat tax reform on hours of work. The fact that the new Russian tax code affected some taxpayers but did not affect the others allows us to use the difference-in-difference (DID) method to identify the effect of tax changes on labor supply. Unlike IKK (2005) who estimate the unconditional mean DID, we employ the DID regression approach that controls for the hourly gross wage rate, non-labor income, other observable labor supply shifters, and constant individual heterogeneity. We allow hours supplied to depend on the two separable (in the log form) components of the net-of-tax wage rate: the gross wage rate and the tax component. In this case, the effect of tax changes on labor supply will be captured through the treatment interaction term.

Thus, we estimate the following DID model:

$$h_{it}^{s} = \beta_1 w_{it} + \beta_2 Z_{it}^{s} + \theta_1 T_i + \theta_2 D_t + \delta\left(T_i \times D_t\right) + \varepsilon_{it}^{s}, \tag{1}$$

where h_{it}^{s} is the log of hours supplied by individual *i* in period *t*, *w* is the log of hourly gross wage rate, Z^{s} is a vector of observable supply shifters (e.g., age, marital status, number of children, non-labor income, etc.), *T* is the treatment indicator (= 1 if in the group that experienced a decline in marginal tax rates and zero otherwise), *D* indicates post-reform period, and ε_{it}^{s} is the error term.[16] The coefficient on the interaction term, δ, tells us how hours supplied by the treatment group differed from that of the control group post reform, *ceteris paribus*, and is our primary parameter of interest. The nice feature of the DID technique is that it controls for most non-tax factors that affect both groups simultaneously. The panel nature of our data allows us to

[16] We use the log-log specification because it improves the statistical properties of our estimates, mitigates the problem of heteroskedasticity, and makes estimates less sensitive to outliers and top coding. Using the log of hours on the left hand side is common in labor supply analysis (e.g., Lee 2001, Pencavel 2002, among others). However, our results hardly change when we use hours instead of the log of hours as dependent variable.

decompose the error term into individual effects, year effects, and the iid "white noise", thus making T_i and D_t redundant while retaining the interaction term.

The above DID model faces several estimation issues. In addition to the common problem of wage endogeneity in labor supply function, the classification of individuals into treatment and control groups poses a serious identification problem. The most obvious choice of the treatment group is all those individuals in the two upper tax brackets whose pre-reform implied gross annual income was greater than 50,000 rubles. Going back to Figure 2, this would be everyone on segments CD and DE of the pre-reform budget constraint. Everyone else would fall into the control group. However, if hours and wages are simultaneously determined, then labor supply via income would influence which tax bracket an individual would fall in, hence, making the assignment into the treatment group based on pre-reform income endogenous.

Using pre-reform income will likely to lead to another significant bias in the estimated treatment effect as a result of possible misclassification of the treatment group due to widespread income underreporting (GMP 2009). High income earners who are facing higher tax rates but choose to underreport their pre-reform income sufficiently will be placed in a lower tax bracket, pay lower taxes, and would be wrongly assigned to the control group when they should be classified as treated. If labor supply among these individuals increases post reform due to lower rates, the misclassification would cause the response of the control group to be overstated and the response of the treatment effect to be understated.[17] This could be one of the reasons why IKK (2005) obtain the negative effect on hours when they define the treatment group based on pre-reform reported earnings.[18]

[17] See GMP (2009) NBER working paper version for a detailed derivation of the evasion bias created if pre-reform income is used to identify treatment and control groups and how using post reform income eliminates this bias.

[18] Besides pre-reform individual reported earnings, IKK (2005) also use pre-reform implied true gross household labor income per working adult to assign individuals to treatment and control groups. While this approach mitigates

To avoid these potential problems, we follow GMP's definition of treatment and control groups and classify individuals as treated if their post-reform implied gross annual income is greater than 50,000 rubles. Using post-reform income accounts for the possible misclassification of individuals due to changes in compliance and thus minimizes the resulting evasion bias. It also addresses the issues of endogenous self-selection into the treatment group. We take advantage of the flat tax rate by exploiting the fact that all taxpayers after reform face the same marginal tax rate regardless of income and irrespective of hours of work. Therefore, the assignment into the treatment and control groups based on post-reform income is not affected by behavioral responses to differential tax rates.

Although the DID method is suitable for estimating the impact of marginal tax rates on hours worked, it still suffers from potential biases. For example, Russia was on the rebound after going through the 1998 financial crisis. If this recovery period affected the treatment group differently than it did the control group, then the DID estimates will be biased unless we fully account for these differences. Additionally, the DID estimates will be biased by any heterogeneity we fail to capture in our model due to unobservables that vary differentially across treatment and control groups over time.

To address these possibilities, we exploit the fact the tax rates changed discontinuously at the threshold of 50,000 rubles under the new tax schedule. Given such a distinct discontinuity in a continuous income variable, the treatment effect can be estimated using individuals that are very similar to each other in every way except for the marginal tax rate they face. In other

---

the tax evasion bias, it is exposed to other biases. For example, the Russian tax system does not allow for family-based tax filing and determines the tax liability based on individual income. Given that many Russian households are large and have more than one earner (56% of working households), individual gross income cannot be correctly inferred from average household net income per working adult. Since the approach assigns everyone in a given household to the same group, households with high and low earners will have at least one misclassified person. The "true labor income" is also likely to be mismeasured as it is based on non-durable expenditures minus pensions and unemployment benefits. As such, it does not account for savings, expenditures on durables, private transfers, capital income and some sources of non-taxable non-labor income such as stipends, social welfare, and child support.

words, we can identify the labor supply effects of the reform by comparing individuals that are just above the income threshold of 50,000 rubles to those individuals that are just below the threshold. These individuals should respond similarly to the economic recovery as well as have similar unobservables not captured by individual fixed effects. This makes it possible to obtain cleaner estimates of the treatment effect than DID (GMP 2009).

3.2. Endogeneity of Wages

The methods discussed above allow us to deal with the endogeneity of the marginal tax rates but does not address wage endogeneity. Wages remain problematic for identification due to the simultaneous determination of equilibrium wages and hours on the labor market. Other potential sources of wage endogeneity include a likely division bias, non-random income underreporting, and measurement error. We address these issues using two different but related approaches. First we derive reduced form wage and labor supply equations from a simultaneous system including labor supply and demand functions. The second method uses demand shifters as instrumental variables for wages in the labor supply regression. In this section, we discuss each approach starting with the reduced form approach. We begin with the following system of structural equations that describe the labor market equilibrium:

$$h_{it}^{s} = \beta_1 w_{it} + \beta_2 Z_{it}^{s} + \delta\left(T_i \times D_t\right) + \varepsilon_{it}^{s} \tag{2}$$

$$h_{it}^{d} = \alpha_1 w_{it} + \alpha_2 Z_{it}^{d} + \varepsilon_{it}^{d} \tag{3}$$

$$h^{s} = h^{d}, \tag{4}$$

where h^d is the log of hours of work demanded by employers, Z^d is a vector of observable labor demand shifters (aggregate and local productivity shocks, technological change, the employer's

share in the labor market, etc.)[19], ε^d is the error term in the demand equation, and the other variables are as defined above. We assume that demand and supply shifters are exogenous and uncorrelated with respective error terms. Imposing the equilibrium condition yields the following reduced form wage equation:

$$w_{it} = -\frac{\beta_2}{\beta_1 - \alpha_1} Z_{it}^s + \frac{\alpha_2}{\beta_1 - \alpha_1} Z_{it}^d - \frac{\delta}{\beta_1 - \alpha_1}(T_i \times D_t) + u_{it}, \tag{5}$$

which we rewrite as:

$$w_{it} = \gamma_1 Z_{it}^s + \gamma_2 Z_{it}^d + \gamma_3 (T_i \times D_t) + u_{it}, \tag{6}$$

where $u_{it} = \frac{\varepsilon_{it}^d - \varepsilon_{it}^s}{\beta_1 - \alpha_1}$. Substituting (5) into (2) gives us the reduced form labor supply equation:

$$h_{it}^s = -\frac{\alpha_1 \beta_2}{\beta_1 - \alpha_1} Z_{it}^s + \frac{\beta_1 \alpha_2}{\beta_1 - \alpha_1} Z_{it}^d - \frac{\alpha_1 \delta}{\beta_1 - \alpha_1}(T_i \times D_t) + \upsilon_{it}, \tag{7}$$

which we rewrite as:

$$h_{it}^s = \pi_1 Z_{it}^s + \pi_2 Z_{it}^d + \pi_3 (T_i \times D_t) + \upsilon_{it}, \tag{8}$$

where $\upsilon_{it} = \frac{\beta_1 \varepsilon_{it}^d - \alpha_1 \varepsilon_{it}^s}{\beta_1 - \alpha_1}$, and π_i are the reduced form estimates.

Equations (7) and (8) imply that $\delta = \pi_3 \left(1 - \frac{\beta_1}{\alpha_1}\right)$. If higher wages increase labor supply and reduce labor demand, then we know that our estimate of $\left(1 - \frac{\beta_1}{\alpha_1}\right)$ will be greater than one. This then implies that the estimated reduced form treatment effect, π_3, will give us a lower bound of the true treatment effect, δ. Therefore, with appropriate demand and supply shifters we can obtain a lower bound estimate of the treatment effect by estimating equation (8) only. In this case, we do not need to worry about the endogeneity of wages.

[19] Demand and supply shifters may partially intersect and be part of both equations. The equation (2) is identified as long as there is at least one exogenous demand shifter that has a nonzero population coefficient in equation (3) and does not affect labor supply.

In principle, it is possible to recover the structural treatment effect by estimating (6) and (8) together and solving the system of equations. However, it is not necessary as long as there are suitable demand shifters that could be used as instruments for wages in equation (2).[20] In this case, equation (2) will be fully identified and give us a consistent and direct estimate of δ. Another advantage of the IV approach is that, unlike the reduced form method, we do not have to worry about the possible independent effect of the reform on labor demand. The IV approach also allows us to estimate the effect of wages on hours worked while the reduced form approach does not.

4. Data

In this section, we introduce the RLMS data that will be used in estimating the outlined empirical model. The RLMS is a comprehensive household survey conducted annually since 1994, except for 1997 and 1999. The panel nature of the data and the fact that we have data before and after the reform are particularly useful in exploiting the "natural experiment" aspect of the reform while accounting for individual heterogeneity. We are also able to capture medium to long-term effects as individuals respond to tax changes. Approximately 10,000 individuals are interviewed each year. We restrict our sample to individuals who were born between 1945 and 1980 (20 to 55 years old in year 2000). This adjustment leaves us with approximately 6,000 observations per year (54% females). These individuals are largely representative of the country population as they come from 38 randomly selected primary sample units (psu or municipalities thereafter) and 32 of 89 constituencies of the Russian Federation (or regions thereafter).

[20] An important requirement of the IV approach is that there should be at least one demand shifter that has a non-zero effect on labor demand but excluded from the labor supply equation. We discuss various instruments in Section 5.

The RLMS has detailed information on multiple measures of hours worked and employment participation. Information on hours of work is available for primary and secondary jobs separately and for supplementary individual economic activities. Figure 3 shows that hours at primary job take up a significant share of all work hours (98%) and the proportion of workers with more than one job has been relatively small and stable for females (7-9%) but declined rapidly for males from 15% to 7% over the ten-year period. Due to the irregular nature of most supplementary individual economic activities, we include those activities in robustness specifications only and focus our labor supply analysis on primary and secondary jobs.

We use actual number of hours worked last month (actual hours, h^a) as well as usual hours of work in a typical week multiplied by 4 weeks (usual hours, h^u). Actual hours tend to be lower and have higher variability than usual hours due to vacations, sick leaves and other reasons (Figures 4). Females typically work 5-6 hours less per week than males but they have larger variability of working hours. Overall, the dispersion of hours falls during 1994-1996 and stays stable afterwards.

Figure 4 also reveals that the trends in hours of work follow the business cycle with a decline until the 1998 financial crisis and a modest increase during the recovery. Compared to the 1990s, the trends in hours of work just before and after the tax reform (2000-2004) were relatively steady and did not experience large fluctuations. Furthermore, if we exclude data from the 1990s, the simple regression of the log of hours worked (actual or usual) on a post-reform dummy controlling for constant individual heterogeneity does not show statistically significant movements in hours of work after 2001 for both males and females. This result does not imply that the reform had no positive effect on the number of hours worked as the aggregate numbers might be hiding significant variation across groups. It could also be the case that other

macroeconomic factors caused aggregate hours to move in a way that cancels the positive tax rate effect. For example, increased share of government transfers and other non-labor sources of income in 2001-2004 may have had a negative income effect on labor supply. It is also possible that increased employment participation among groups with typically lower hours (females, young cohorts, and college graduates) may have decreased the average aggregate hours. Because of these and other confounding factors, it is important that a proper identification strategy be used to determine the effect of the reform on labor supply.

Besides hours of work, we also examine the employment participation response to tax changes. Panel A of Figure 5 depicts a very high employment-to-population ratio for both males and females, which is commonly observed in transition economies. We see that the employment-to-population ratio has followed a pronounced U-shaped path in the 1990s, but it stabilized around year 2000 and stayed at 84-86 percent for males and 78-80 percent for females in subsequent years.[21] Again, as with hours, we do not observe a noticeable shift in employment participation following the 2001 tax reform.

Individual labor earnings are another essential variable in our empirical model. Respondents are asked to report two measures of labor earnings: monetary after-tax earnings received last month (actual earnings) and average monthly after-tax earnings in the last 12 months (contractual earnings). These measures of labor earnings are used with the corresponding monthly hours to calculate actual and contractual hourly net wages, w^a and w^c, depicted in panel *B* of Figure 5. We observe that actual earnings received last month are much more variable than contractual earnings, which is partly due to widespread wage arrears and irregular wage payments during 1994-1998. Because gross declared earnings are not available in

[21] See Gorodnichenko, Peter, and Stolyarov (2009) for a more detailed discussion of the trends in employment and hours.

RLMS, we compute them based on after-tax reported earnings and the PIT rate schedule with deductions. Appendix A1 provides details of computations.

In addition to providing the overall macroeconomic picture, the above trends in hours, employment, and wages are also helpful in determining the appropriate time span for our analysis. For reasons of high volatility in both hours and wages and the lack of measures on usual hours and contractual wages, we exclude the 1994-1996 period from our analysis. In baseline specifications, we also leave out year 1998 to avoid the potential influence of the financial crisis on the treatment effect estimates, and year 2005 to make our labor supply response estimates comparable with the consumption-based estimates of total earned income response to tax changes in GMP (2009). Thus, our baseline specification is estimated for the 2000-2004 sample period. However, we do include 1998 and 2005 in supplementary specifications to test for robustness of our results.

A standard labor supply model requires several other variables, including non-labor family income, age, years of schooling, marital status, and number of children, among others. We provide more details on each variable in Appendix Table A1 and their summary statistics by gender and treatment status in Table 1. The treatment status is determined based on post-reform actual labor earnings from primary and secondary jobs.[22] From Table 1 we observe that treated males are relatively younger, more educated, and more likely to be married and have children than the control group. Treated males also have greater non-labor income, earn higher wages, likely to have more than one job, and work longer hours compared to males in the control group. With the exception of age, children, and marital status, similar observations are made for treated females relative to females in the control group. Unlike their male counterparts, treated females

[22] The new PIT schedule mainly affected labor income. Most forms of capital income are taxed at different rates while private and public transfers remain non-taxable. The share of taxable non-labor income is less than 1% in our data. In Section 5, we show the sensitivity of our estimates to alternative definitions of the treatment group.

are less likely to be married and have children and are the same age as females in the control group. In addition to mean hours, we also plot the distribution of actual hours by treatment status in Figure 6. While there is no visual evidence that the distribution of hours changed after reform for either group of females, we do observe a slight shift in the distribution of male hours to the left for the control group and to the right for the treatment group. From the plotted distributions, it is not clear whether labor supply response varies along the percentiles of hour distribution. To reach more definite answer, we perform the quantile regression analysis that will be discussed in section 5.6.

5. Results

5.1. DID Baseline Model

We first present the estimates of the baseline labor supply model for males and females separately. We estimate four equations: equation (1) with year effects by OLS; equation (2) with individual and year fixed effects by OLS; equation (2) with individual and year fixed effects by IV; and reduced-form specification as in equation (8) with individual and year fixed effects by OLS. The baseline model includes the log of actual hours worked in the last 30 days on the left-hand side and controls for the log of hourly gross wage rate w^c, age, years of schooling, marital status, number of children, the log of non-labor family income, and a dummy variable indicating if labor is the only source of family income.[23] All constant local and family characteristics are controlled by default when individual fixed effects are included. Treatment and control groups are defined using positive actual labor earnings at primary and secondary jobs. All estimates report Arellano (1987) robust standard errors to correct for arbitrary serial correlation.

[23] The log of non-labor income is calculated as log (1+non-labor income). Therefore, individuals whose sole source of income is from labor would have a value of zero. To control for the kink at zero value, we include a dummy variable indicating if labor is the only source of family income.

Both IV and reduced-form approaches require labor demand shifters as exclusion restrictions. Unlike most household surveys, the RLMS asks respondents about the characteristics of their employers. One of those characteristics, firm employment size, might be argued to influence the demand for labor. We can appeal to both the standard monopsony model and the size-wage premium literature (e.g., Brown and Medoff 1989) to justify the choice of firm size as a labor demand shifter. Another employer characteristic asked in the RLMS is whether the firm is owned by foreign entities. We can use this variable to calculate the local (psu) share of foreign firms in total employment and refer to the large body of FDI literature that examines the effect of foreign firm concentration on domestic productivity (and thus on the demand for product and labor) via spillovers and crowding out (e.g., Aitken and Harrison 1999, Haddad 1993, and Javorcik 2004).[24] Finally, we turn to external data sources for some regional labor demand shifters that reflect changes in aggregate employment and productivity. Consequently, our vector of labor demand shifters include 9 categories of firm size, the local employment share of foreign firms, the annual difference in the log of regional employment, and the annual difference in the log of regional value added per capita. Summary statistics for the demand shifters are presented in Table A2. We make sure that these instruments are jointly statistically significant in the first stage and pass the over-identification test.[25]

We report the DID estimates of the baseline model for males and females in Tables 2 and 3, respectively.[26] As indicated earlier, our primary interest is in the estimated treatment effect. The OLS result for males yields a treatment effect of 0.073 log points, which indicate that males'

[24] Using Russian firm-level data, Sabirianova, Svejnar, and Terrell (2005) find that the share of foreign firms in employment has a negative effect on the productivity of domestic firms and positive effect on other foreign firms.
[25] The firm size variable was subsequently dropped from female specifications because it failed the IV validity test. Specifically, no statistically significant relationship is found between the firm size and female wages.
[26] Hours of work is found to be concave in age for both males and females. With the exception of the reduced form specification, these results are statistically significant. We find no statistically significant effect of marital status or years of schooling on hours worked except in the OLS specification where married males (females) work more (fewer) hours. The qualitative aspects of these results remain the same in all of our specifications.

actual hours worked increased by approximately 7.3% as a result of the lower tax rates implemented in the tax reform. The estimate falls by 2 percentage points to 5.2% when we control for time-invariant individual heterogeneity, but remains statistically significant at the 1% level. When we use demand shifters as instruments for wages in the third column, the treatment effect estimate is of the same magnitude (5.3%) and indicates that the reform had a statistically significant positive effect on male labor hours. We check that the instruments are jointly statistically significant in the first stage at the 1% level and that the over-identification test does not reject the null that the instruments are exogenous (Hansen J-Statistic p-value is 0.329). We also use the demand shifters to estimate the treatment effect in the reduced form labor supply model. The estimated effect of 5.1% is slightly smaller than that obtained using IV and is statistically significant at the 1% level. The fact that the reduced form estimate is lower than the IV one is consistent with our earlier arguments that the reduced form estimates are a lower bound to the true impact of the reform. The results in Table 3 indicate that the reform had no effect on female hours of work; all estimates are statistically and economically insignificant.

5.2. Sensitivity Analysis

The following section shows how the above estimates of the treatment effect are sensitive to alternative measures of hours of work (actual vs. usual hours), different coverage of jobs (hours at primary job, primary and secondary job, and all jobs), definitions of the treatment and control groups (actual vs. contractual labor earnings), and different samples (restricted sample and sample with additional years 1998 and 2005). We report the treatment effect estimates from the two specifications: equation (2) with individual and year fixed effects by IV (Table 4) and equation (8) with individual and year fixed effects by OLS (Table A3). Each cell in Tables 4 and A3 presents the estimated coefficient on the interaction term between the treatment group (upper

brackets) and a post-reform dummy using different measures of hours, earnings, and samples.[27] All specifications include the same set of variables as in Tables 2 and 3.

The results in Table 4 mostly confirm the results discussed in Tables 2 and 3. From panel A, we find that the treatment effect for males ranges from 4.5% (primary job only) to 6.7% (all jobs) when we use actual hours as our dependent variable. This implies that actual hours increased in all jobs, including secondary employment and individual economic activities, as a result of the reform. The same pattern is observed for usual hours, although the estimates are smaller: 3.1% for primary job and 4.8% for all jobs. This result is not surprising since usual hours of work are subject to less erratic variation over time.

Panel B of Table 4 reports additional robustness checks based on hours worked at primary and secondary job. First, we observe that including 1998 or 2005 in the estimation sample does not have any significant effect on the estimates. We also restrict the sample to individuals whose implied gross annual earnings are between 4,800 and 100,000 rubles. We use the restricted sample to remove the influence of standard deduction and eliminate the differential effect of the regressive unified social tax paid by employers. For each worker in the restricted sample, employers should pay the same 35.6% rate of the unified social tax after 2000. Another benefit of the restricted sample is that the treatment group would have faced the same 21% rate if the pre-reform tax scale had remained after 2000 (the next 31% bracket begins with 150,000 rubles). In another robustness check, we add interactions between district dummies (at the *psu* level) and year dummies to control for time-varying local shocks. Finally, we use contractual labor earnings instead of actual earnings in defining the treatment group. In each case the estimated treatment effect for actual hours remains in the range of 4.6–6.1% and is statistically

[27] We report Arellano (1987) robust standard errors to correct for arbitrary serial correlation. The IV estimations in Table 4 are performed using the Stata 10 module *xtivreg2* (Schaffer 2007). The conventional treatment of outliers is also applied.

significant. The estimates for usual hours of work remain marginally smaller than those for actual hours but maintain their statistical significance and quantitative magnitude (4–5.6%). Table 4 also reports estimates of the treatment effect for females. We find no effect on actual or usual hours of work regardless of specification.

Additional supporting evidence of the pattern of change we have observed so far is presented in Table A3, where the reduced form specification is used. There we observe that the reform had a positive and statistically significant effect on both actual and usual hours of work for males and no effect on females. The estimates for males are modestly smaller in size than those reported in Table 4 (3–5%) but remain statistically significant at the 5% level. On the other hand, the estimates for females are statistically insignificant regardless of specification.

These results are different from previous estimates by IKK (2005). While they find no evidence that the reform affected number of hours worked, we find that the reform affected male hours positively and had little or no effect on female hours.[28]

Because the reform affected two upper brackets of the pre-reform tax schedule, we also examine the effect of tax reform on individuals in each bracket (21% and 31%). The purpose of this exercise is to see if the strength of labor supply response varies with income. The results from this specification (both IV and reduced form) are reported in Table A4. They confirm the previous finding of a significant positive effect on males and no effect on females. More interesting though, is the fact that individuals in both brackets responded similarly to tax changes. Male earners in the top bracket increased their hours more than those in the middle

[28] We obtain qualitatively similar results to that reported in IKK (2005) when we replicate their approach. In particular, we estimate our OLS specification in Tables 2 and 3 with no distinction between males and females, no explanatory variables (except for the post reform dummy, the treatment dummy, and their interaction), and only two years (2000 and 2001). When we do this, the treatment effect becomes statistically insignificant as in their paper. When we follow the same approach above but stratify by gender, we find a statistically significant positive effect for males and no effect for females. We also get a statistically significant positive effect if we include additional years and/or explanatory variables. It is clear that ignoring these dimensions of the problem contributed to the finding reported by IKK (2005). These results are not reported but can be obtained upon request from the authors.

bracket, but the difference between the two brackets is not statistically significant in 3 of 4 specifications.

To further assess the validity of our results, we test if there is labor supply response within the control group. We divide the control group into two groups of high and low earners at three different points: standard deduction threshold (4,800 rubles), the highest annual income at which standard deduction can be applied (20,000 rubles), and the random point of 40,000 rubles. In all three cases, we do not find statistically significant differences in labor supply response between high and low earners of the control group after the reform. To save space, the results are not reported but available upon request.

5.3. Weighted Difference-in-Difference

Our next set of estimates employs the above DID method but weights each individual observation by the distance of reported earnings from the threshold of 50,000 rubles. GMP (2009) use this weighted difference-in-difference (WDID) approach in estimating the tax evasion response to tax changes. We apply the same method to identify the labor supply response at the point of discontinuity around the income threshold. In this case, the treatment effect is less likely to be confounded with other factors as individuals are drawn from the same portion of income distribution and likely to have similar unobservables. In a way, the WDID estimation is analogous to the regression discontinuity (RD) design.

Following GMP, we write the estimated model as:

$$\sum_{i=1}^{n} \psi_{ip} \left(h_{it}^{s} - \beta_1 w_{it} - \beta_2 Z_{it}^{s} - \delta \left(T_i \times D_t \right) - \varepsilon_{it}^{s} \right)^2, \tag{9}$$

where ψ_{ip} is the weight assigned to individual i in post reform period p. The weights are a decreasing function of the distance between the post-reform earnings and pre-reform threshold.

They are calculated as $K\left(\frac{lnY_{it}^{R}-lnY_{t}^{\#}}{b}\right)\Big/\sum_{j=1}^{n}K\left(\frac{lnY_{jt}^{R}-lnY_{t}^{\#}}{b}\right)$, where $K(*)$ is the Gaussian kernel density, $b\approx0.17$ is the optimal Gaussian "plug-in" bandwidth, Y^R is post-reform reported earnings and $Y^{\#}$ is the pre-reform threshold. The use of post-reform earnings in defining the treatment and control groups is especially critical for the RD-type approaches requiring that the income variable has no discontinuous jumps at the threshold. This requirement is violated if we use pre-reform income due to the tax evasion response, which leads to the clustering of reported earnings just below the threshold prior to the reform. Such clustering undermines the fundamental assumption on which the RD approach is based, thus leading to biased estimates.

Table 5 reports the WDID estimates with individual fixed effects. Again, we use demand shifters as instrumental variables for endogenous wages. The estimated treatment effects are qualitatively similar to those obtained using the unweighted DID technique (Table 4). Specifically, the range of the WDID estimates for males overlap with that of DID: 4–6% for actual hours and 3–5% for usual hours.[29] The results for females in columns 4 and 5 confirm our conclusion up to this point that the reform did not have a statistically significant effect on hours supplied by females.

5.4. Elasticity of Labor Supply

In this section we discuss the size of labor supply elasticities with respect to marginal tax rates implied by the estimated treatment effects for hours of work. This exercise is especially important for researchers and policy makers interested in understanding the implications of future policy changes. Additionally, the estimates can be used in computable general

[29] The WDID estimate of the treatment effect is fairly insensitive to any bandwidth larger than 0.1. If the bandwidth falls below 0.1 (which implies a steep decrease in the weighting function), the treatment effect maintains its magnitude but becomes statistically insignificant due to the insufficient number of observations around the income threshold.

equilibrium models and other macroeconomic models geared towards quantifying the impact of economic policies on the economy.

The elasticity of working hours with respect to tax rates is calculated as $\in_{h\tau} = \frac{\partial h}{h}\frac{\tau}{\partial \tau} = \delta \frac{\tau}{\partial \tau} = \delta \frac{0.21}{(-0.08)} = -2.63\delta$, where δ is the estimated treatment effect, τ is the marginal tax rate, and $\partial \tau = 0.13 - 0.21 = -0.08$ is the change in tax rate. For example, the labor supply elasticity derived from the baseline IV estimate for males (Table 2) is $-2.63 \times 0.053 \approx -0.14$. This implies that a 1% reduction in the marginal tax rate increases male hours of work by 0.14%, *ceteris paribus*. The elasticity estimates based on alternative samples, specifications, and methods (Tables 4 and 5) fall in a fairly small band of approximately -0.08 to -0.18. The labor supply elasticity for females is not statistically significant from zero.[30]

Given that the share of females in the treatment group is about 40%, the population-averaged elasticity of the labor supply response to tax changes for the baseline specification becomes $(-0.14 \times 0.6 + 0 \times 0.4) \approx -0.084$. Although our elasticity estimates point to a very weak response of hours to changes in tax rate, they are not implausible (Meghir and Phillips 2008). For example, Immervoll *et al* (2007) used elasticity estimates in the range 0.1–0.2 for their simulations of welfare reforms on income distribution and efficiency in Europe. Kumar (2008) cites several studies that find little or no changes in labor supply following the 1986 U.S. tax reform. Our estimates are consistent with the small elasticities of total earned income with respect to tax rate (about –0.05) obtained by GMP (2009) based on the consumption response to the Russian flat tax reform.

[30] These labor supply elasticities are closer to the compensated Frisch-type elasticities. The Frisch elasticity requires the marginal utility of wealth to be constant. In our case, individual fixed effects account for all time-invariant factors of labor supply, including the marginal utility of wealth that is unlikely to change over the short time span of our data.

The small magnitude of labor supply response implies that it is unlikely that the observed spike in the after-reform PIT revenue could be explained by increased hours of work due to the flat tax rate. Suppose that PIT revenue could be roughly approximated as $R = W \times h \times t$, where W is the observed gross hourly wage rate. Given that the share of the treated males in total working population is about a third, a post-reform increase in aggregate hours of work should be $5.3\% \times 0.33 \approx 1.7\%$ over the 4-year period. Consequently, increased hours of work in response to tax changes cannot account for more than a 1.7% increase in overall PIT revenue.

5.5. Extensive Margins

The above estimates of labor supply response are conditional on being employed and thus do not take into account the possible impact of the reform on employment participation. It is now widely accepted that responses along extensive margins (participation) are just as important as intensive margin responses (hours) (Blundell and MaCurdy 1999 and Saez 2002). However, the estimation of responses along extensive margins is complicated by the fact that we do not observe labor earnings for non-participants and hence cannot assign these individuals to treatment or control groups. One way of dealing with this problem is to assign non-participants to treatment and control groups based on their post-reform predicted earnings.[31] Then for those individuals that did not have a job in 2000, we can estimate the effect of tax changes on the probability of finding a job after reform.

Suppose the probability of finding a job p_{it} conditional on observables X_{it} is defined as $p_{it} = P\left(E_{it} = 1 \mid E_{i,2000} = 0, \mid X_{it}\right)$, where E_{it} is the indicator of employment participation for

[31] Labor earnings for working-age non-participants are predicted using gender, age, age squared, years of schooling, marital status, number of children, non-labor household income, regional labor demand shifters, and 38 dummies for municipalities.

individual i in year t. Then, we can estimate the following DID linear probability model with fixed effects:

$$E_{it} = \lambda_1 Z_{it}^s + \lambda_2 Z_{it}^d + \lambda_3 (T_i \times D_t) + \mu_i + \theta_t + \varepsilon_{it}, \tag{10}$$

where E_{it} is a dummy indicating whether a respondent worked at least one hour at primary job in the last 30 days, Z^s and Z^d are vectors of observable labor supply and demand shifters, respectively, T is the treatment indicator, D indicates post-reform period, μ_i and θ_t are individual and time fixed effects, and ε_{it} is the *iid* error term. The estimated model has the same set of explanatory variables as in Table A3.

Table 6 reports the estimates for each of these participation measures using either actual labor earnings (columns 1 and 3) or contractual earnings (columns 2 and 4) in defining the treatment group. As before, we estimate the models for males and females separately. The results show that the reform had a statistically significant effect on the probability of job finding. In fact, the probability of making the transition to work increases by approximately 0.08–0.14 for both treated males and females regardless of which labor earnings we use to determine treatment. We therefore find that treated individuals are more likely to transition into employment following the reform relative to individuals in the control group.

In a similar vein, we estimate the effect of reform on the probability of making the transition from one job to multiple jobs. Our discussion in Section 2 revealed that changes in the Russian tax code may have created incentives to have (or report) additional formal jobs; yet the total effect of tax changes on multiple job holding from the theoretical point of view is ambiguous. The results in Table 6 indicate that the probability of having additional jobs increased by 0.018–0.023 for females as a result of the reform. There was no significant effect for males.

5.6. More on Female Labor Supply

The finding that female hours did not respond to the reform is an unusual result when one considers the vast international literature on labor supply (Meghir and Phillips 2008). Our attempts to search for an explanation of this result by interacting the treatment terms with various individual and household characteristics turned out to be mostly unsuccessful. We find no evidence that the female treatment effect varies by marital status, number and age of children, household composition, respondent's age, schooling, location, firm size, job sector, and other variables (the results are not reported to preserve space). These results, for example, rule out the explanation that females who fulfill additional household functions (e.g., having pre-school children) may find it more difficult to adjust working hours.

At the same time, we know from the previous section that females did respond in terms of employment participation. They also have a higher probability of switching to multiple jobs after the reform. The fact that females responded along extensive, but not intensive, margins can be due to the lack of hour flexibility at the primary place of employment. If this is the case, then we should probably expect a larger response among females working part-time as their hours tend to be more flexible. To test this hypothesis, we estimate equation (1) at different points of the hour distribution by using quantile regressions and report the estimated treatment effect at each percentile in Figure 7. The findings confirm our expectations. We observe that the female labor supply response is relatively large and positive in the bottom and very top parts of the distribution where hours tend to be more flexible. The response among females in the middle of the distribution is flat at around zero. This is contrary to the pattern observed for males. The male labor supply response is highly volatile in the lower tail (due to small number of males in part-time employment) but increases steadily along the distribution from the 30th percentile

upward. The treatment effect among males is large and positive in the second half of the distribution. As a robustness check, we also interacted all treatment and time terms in equation (1) with the part-time status (<120 hours worked last month) and find that part-time females tend to have a stronger and positive response to tax changes than females working full time. Thus, the zero effect found for females in the mean regressions is likely to be due to the rigidity of hours in the middle of distribution.

6. Conclusions

The Russian tax reform represents one of the major tax reforms of the 21st century. Despite its significant popularity, very little in the way of empirical analysis has been done to test its impact on real responses. We attempt to bridge this gap in the literature by estimating the impact of the reform on several margins of labor supply including hours of work. The estimated hours responses are used to calculate implied elasticities that can be used in general equilibrium models to predict the possible impact of future policy changes.

Using micro level data on individuals from the RLMS, we implement DID and RD-type methodologies to identify the treatment effect of the reform on actual and usual hours of work. The nature of the reform allows us to address the endogeneity of marginal tax rates while the endogeneity of wages is addressed using labor demand shifters as instrumental variables. We estimate labor supply functions separately for males and females to account for the fact the females respond differently than males to labor market incentives.

Regardless of the estimation technique used, we find that the reform had a positive impact on male hours of work with implied elasticities - with respect to tax rates - ranging from -0.08 to -0.18. Although relatively small, these estimates are in line with existing male labor supply elasticities. The more surprising result is that the reform had no statistically significant

effect on female average hours of work. Further analysis reveals that while this result is true on average, there is significant variability in the treatment effect at different percentiles of the corresponding hour distribution. In particular, we find a positive labor supply response among part-time female workers and also at the very top of the female hour distribution where hours tend to me more flexible. However, female hours are fairly rigid in the middle of the respective hour distribution.

We also find evidence that individuals responded to the flat tax rate along the extensive margins of labor supply. The probability of finding a job increased for both males and females, and more females transitioned from one job to multiple jobs after the reform. Our results indicate that the aggregate estimates of labor supply effects are likely to conceal considerable variation in the response across individuals and various margins.

One of the major questions surrounding the Russian flat tax reform is whether or not the increase in PIT revenues following the reform can be credibly attributed to the change in marginal tax rates. Although our empirical results show that the number of hours worked by males increased as a result of the lower PIT rates, there is no evidence that the increased PIT revenues can be explained by this response. In fact, the revenue gains from the increase in hours of work are not sufficient to make up for the lost revenue resulting from lower PIT rates.

Our results along with a well established labor supply literature make it clear that aggregate hours of work may very well be the smallest source of response to tax changes. In the case of Russia, for example, the response in compliance proved more important than aggregate hours for the revenue response. This implies that countries with high levels of non-compliance are more likely to benefit from flattening their tax schedules than those where compliance is

relatively high. It is therefore important that policy makers consider carefully the institutional context of their respective countries before adopting similar tax policies.

References

Aitken, B. J. and Harrison, A. E. (1999). 'Do Domestic Firms Benefit from Foreign Direct Investment? Evidence from Venezuela.' *American Economic Review* 89(3), 605-618.

Altonji, J. G. and Paxson, C. H. (1992). "Labor Supply, Hours Constraints, and Job Mobility." *Journal of Human Resources* 27(2), 256-278.

Arellano, M. (1987). "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics* 49(4), 431–434.

Boeri, T. and Terrell, K. (2002). 'Institutional Determinants of Labor Reallocation in Transition.' *Journal of Economic Perspectives* 16(1), 51-76.

Blundell, R. and MaCurdy, T.E. (1999). 'Labor Supply: A Review of Alternative Approaches.' In: O. Ashenfelter and D. Card, Editors, *Handbook of Labor Economics*, Vol. 3A, Elsevier, Amsterdam.

Brown, C. and Medoff, J. (1989). 'The Employer Size-Wage Effect.' *Journal of Political Economy* 97(5), 1027-59.

Conway, K. S. (1997). 'Labor Supply, Taxes, and Government Spending: A Microeconometric Analysis.' *Review of Economics and Statistics* 79(1), 50-67.

Dhami, S. and Al-Nowaihi, A. (2007). 'Why Do People Pay Taxes? Prospect Theory Versus Expected Utility Theory,' *Journal of Economic Behavior and Organization* 64(1), 171-192.

Eissa, N. and Liebman, J. B. (1996). 'Labor Supply Response to the Earned Income Tax Credit.' *Quarterly Journal of Economics* 111(2), 605-37.

Feldstein, M. (1995). 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act.' *Journal of Political Economy* 103(3), 551-572.

Gaddy, C. G. and Gale, W. G. (2005). 'Demythologizing the Russian Flat Tax.' *Tax Notes International*, March 14, 983-988.

Gorodnichenko, Y., Martinez-Vazquez, J. and Sabirianova Peter, K. (2009). 'Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia.' *Journal of Political economy* 117(3), 504-554.

Gorodnichenko, Y., Sabirianova Peter, K. and Stolyarov, D. (2009). 'Inequality and Volatility Moderation in Russia: Evidence from Micro-Level Panel Data on Consumption and Income,' NBER Working Paper, No. 15080.

Haddad, M. and Harrison, A. (1993). 'Are There Positive Spillovers from Direct Foreign Investment? Evidence from Panel Data for Morocco.' *Journal of Development Economics* 42(1), 51-74.

Hall, R. E. and Rabushka, A. (2007). *The Flat Tax.* Stanford: Hoover Institution Press.

Immervoll, H., Kleven, H. J., Kreiner, C. T. and Saez, E. (2007). 'Welfare Reform in European Countries: A Microsimulation Analysis.' *Economic Journal* 117(516), 1-44.

Ivanova, A., Keen, M. and Klemm, A. (2005). 'The Russian 'Flat Tax' Reform.' *Economic Policy* 20(43), 397-435.

Javorcik, B. S. (2004). 'Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages.' *American Economic Review* 94(3), 605-627.

Kumar, A. (2008). 'Labor Supply, Deadweight Loss and Tax Reform Act of 1986: A Nonparametric Evaluation Using Panel Data.' *Journal of Public Economics* 92(1-2), 236-53.

Lee, Chul-In. (2001). 'Finite Sample Bias in IV Estimation of Intertemporal Labor Supply Models: Is the Intertemporal Substitution Elasticity Really Small?' *Review of Economics and Statistics* 83(4), 638-646.

Lemieux, T., Fortin, B. and Frechette, P. (1994). 'The Effect of Taxes on Labor Supply in the Underground Economy.' *American Economic Review* 84(1), 231-54.

Martinez-Vazquez J., Rider M., Qibthiyyah R. and Wallace S. (2006). "Who Bears the Burden of Taxes on Labor Income in Russia?" International Studies Program Working Paper Series, No. 06-21.

Matveenko, V. and Saveliev, P. (2005). 'Labor Supply in Russia: Studying the Role of Outside Options of the Employed.' EERC Working Paper Series, No. 00-215e.

Meghir, C. and Phillips, D. (2008) 'Labour Supply and Taxes.' IZA Working Paper, No. 3405.

Pencavel, J. (2002). 'A Cohort Analysis of the Association between Work Hours and Wages among Men.' *Journal of Human Resources* 37(2), 251-274.

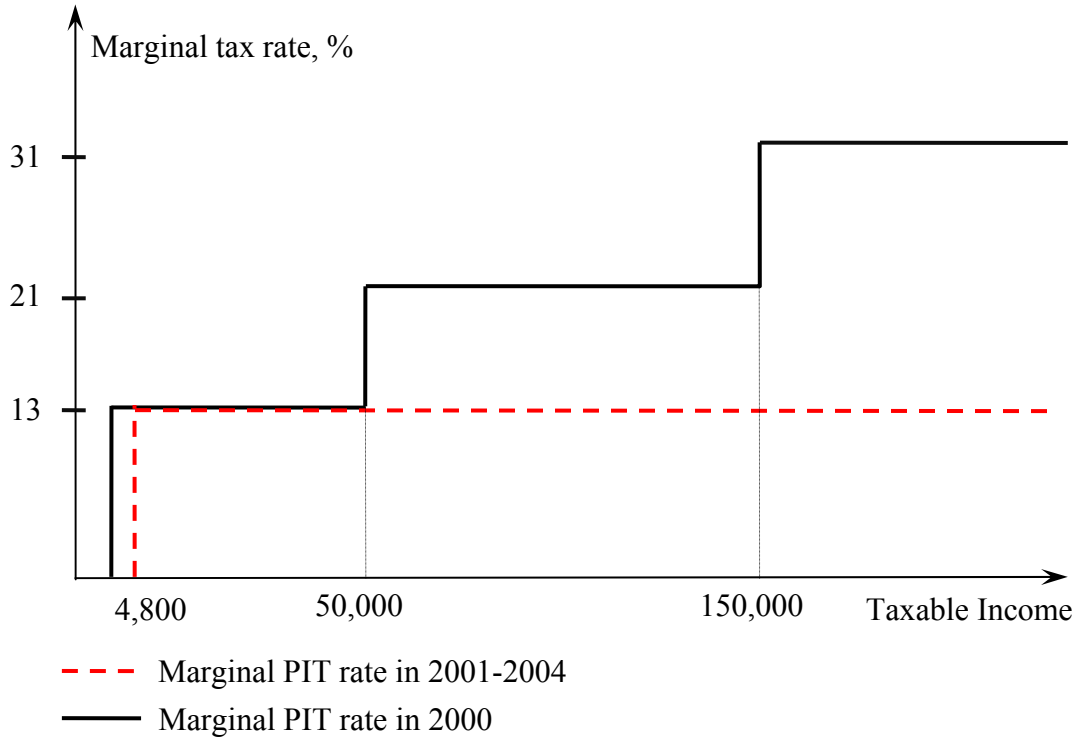
Sabirianova, K., Svejnar, J. and Terrell, K. (2005). 'Distance to the Efficiency Frontier and Foreign Direct Investment Spillovers.' *Journal of the European Economic Association* 3(2-3), 576-586.

Saez, E. (2002). 'Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses.' *Quarterly Journal of Economics* 117(3), 1039-1073.

Schaffer, M.E. (2007). 'XTIVREG2: Stata Module to Perform Extended IV/2SLS, GMM and AC/HAC, LIML and k-class Regression for Panel Data Models. http://ideas.repec.org/c/boc/bocode/s456501.html

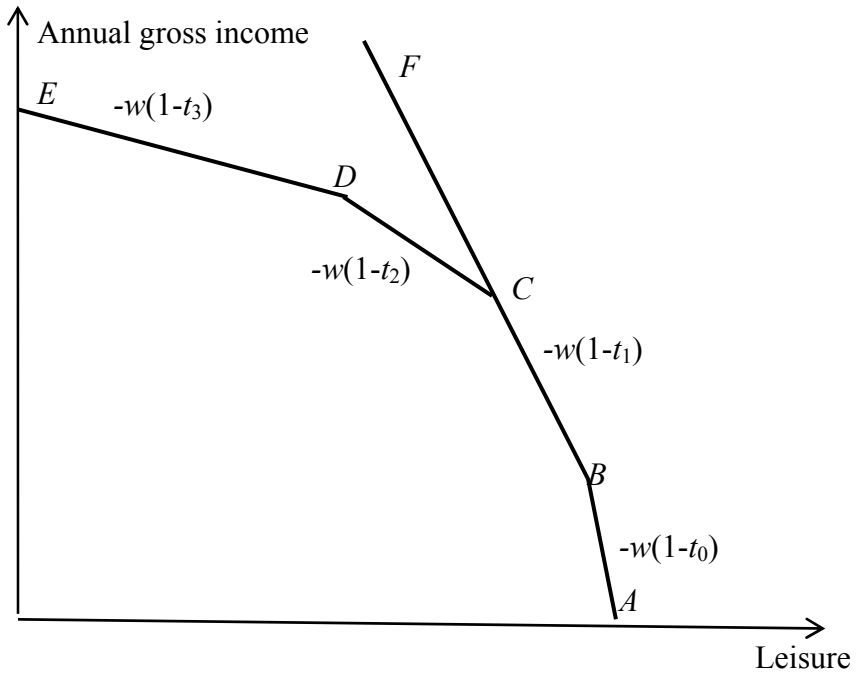
Slemrod, J. (1992). 'Do Taxes Matter? Lessons from the 1980's.' *American Economic Review* 82(2), 250-56.

Figure 1: Marginal Personal Income Tax Rate Before and After Reform



Notes: Taxable income is annual and in rubles. The 2000 marginal PIT rates include 1 percent individual contribution to the pension fund. Standard deductions were 3,168 rubles in 2000, but they increased to 4,800 rubles in 2001.

Figure 2: Budget Constraint Implied by the Tax Schedule Before and After Reform

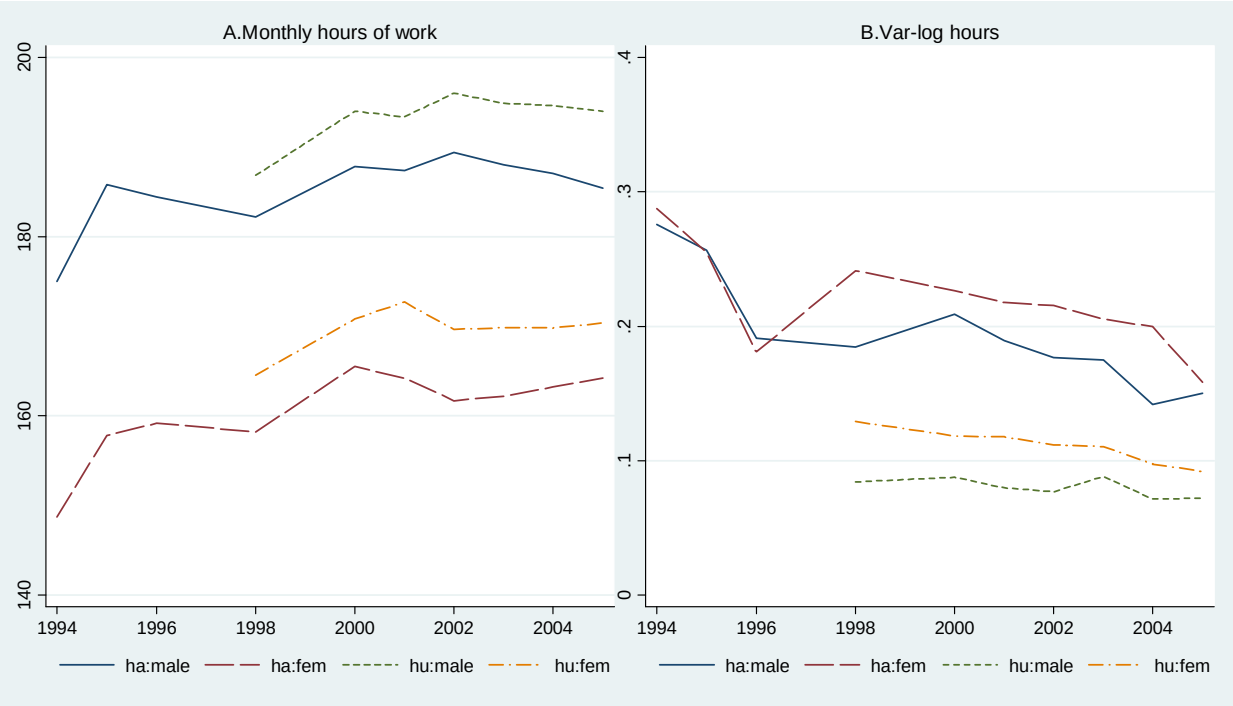


Notes: The slope of each segment is given by $-w(1-t_i)$, where w is the gross wage rate and t_i is the marginal tax rate defined as t_0=0.01, t_1=0.13, t_2=0.21, t_3=0.31.

Figure 3: Trends in Labor Supply

Notes: Panel A shows trends in actual hours worked last month at primary job (*pj*), primary and secondary jobs (*ps*), and all jobs. Sample is restricted to individuals who were between 20-55 years old in the year 2000.

Figure 4: Actual and Usual Hours of Work by Gender

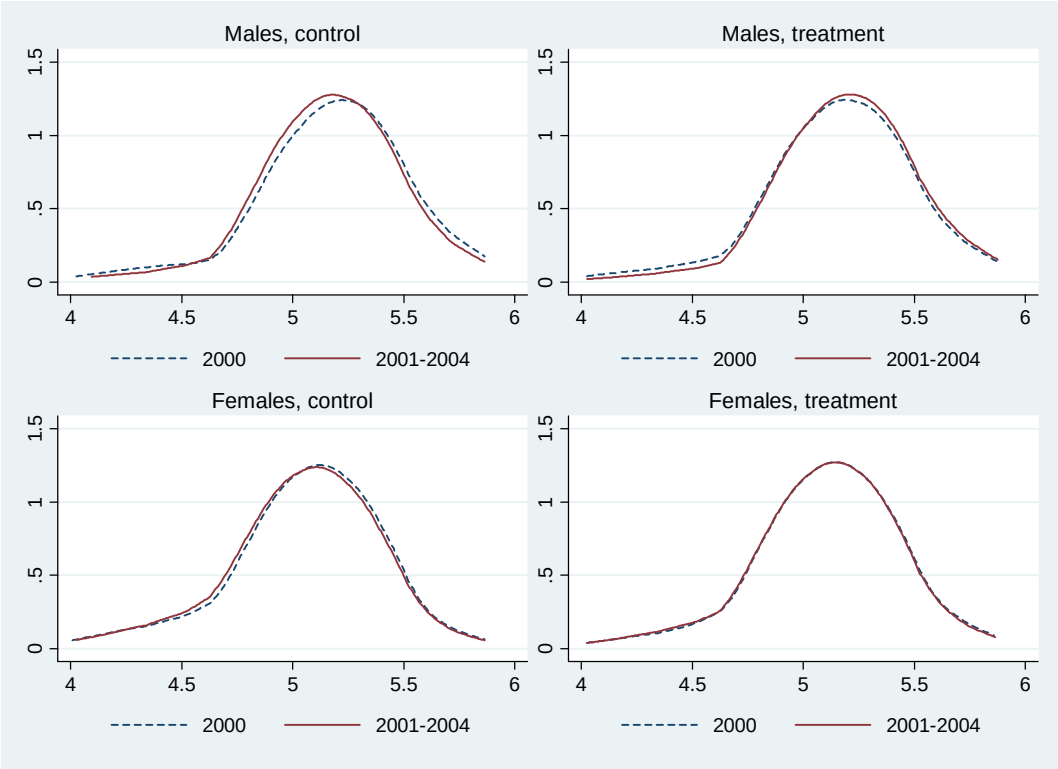


Notes: *ha* = actual hours worked last month; *hu* = usual hours of work per month. Hours in both panels are totals for primary and secondary jobs. Sample is restricted to individuals who were between 20-55 years old in the year 2000.

Figure 5: Trends in Employment and Wages

Notes: *wa* = hourly net wage rate based on earnings received last month; *wc* = contractual hourly net wage rate. All wages are after-tax and deflated with national monthly CPI. Wages in panel B are totals for primary and secondary jobs. Sample is restricted to individuals who were between 20-55 years old in the year 2000.

Figure 6: Distribution of Hours Before and After Reform



Notes: The log of hours is plotted on horizontal axes and truncated at 4. Sample is restricted to individuals who were between 20-55 years old in the year 2000.

Figure 7: Quantile Estimates of the Treatment Effect

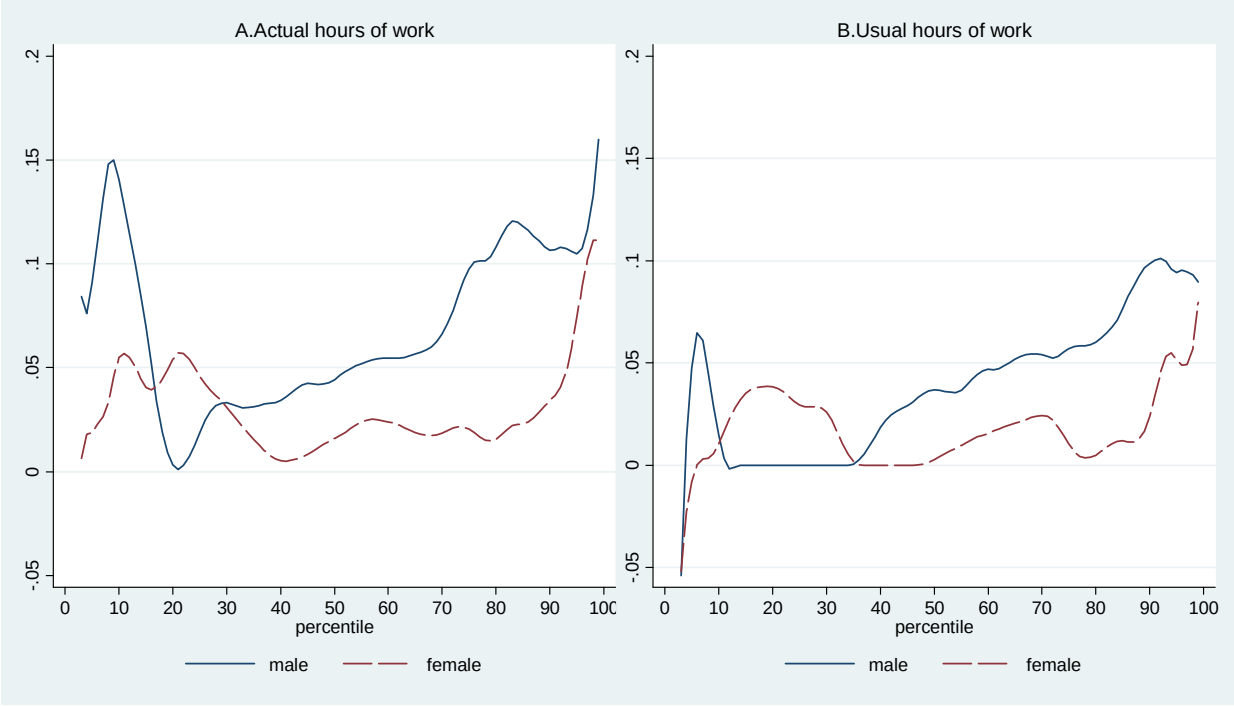


Table 1: Summary Statistics

	Males			Females		
	Control Group	**Treated Group**	**All**	**Control Group**	**Treated Group**	**All**
Age	39.487	37.502	38.058	39.273	39.793	39.524
	(10.314)	(9.635)	(9.870)	(9.806)	(9.579)	(9.700)
Years of schooling	11.095	12.133	11.842	12.072	13.040	12.540
	(2.018)	(2.177)	(2.184)	(1.973)	(2.011)	(2.049)
Married	0.833	0.862	0.854	0.724	0.679	0.702
	(0.373)	(0.345)	(0.353)	(0.447)	(0.467)	(0.457)
1 or 2 children in HH	0.419	0.484	0.466	0.460	0.401	0.431
	(0.494)	(0.500)	(0.499)	(0.498)	(0.490)	(0.495)
3 or more children in HH	0.025	0.010	0.014	0.012	0.006	0.009
	(0.156)	(0.098)	(0.117)	(0.110)	(0.076)	(0.095)
Child <7 years old in HH	0.285	0.312	0.304	0.252	0.204	0.229
	(0.451)	(0.463)	(0.460)	(0.434)	(0.403)	(0.420)
Non-labor family income last month (rubles)	2277	3598	3229	3166	4689	3903
	(2368)	(4923)	(4401)	(3642)	(5971)	(4967)
Labor only source of Income (dummy)	0.066	0.100	0.091	0.053	0.089	0.070
	(0.248)	(0.300)	(0.287)	(0.224)	(0.285)	(0.256)
Contractual gross wage rate (rubles)	9.057	26.216	21.413	8.722	22.765	15.514
	(6.781)	(26.300)	(23.880)	(6.249)	(19.957)	(16.187)
Actual hours last month	187.163	190.631	189.660	161.362	170.706	165.881
	(54.043)	(55.494)	(55.111)	(49.298)	(48.560)	(49.162)
Usual hours per month	189.150	194.306	192.863	167.998	175.894	171.817
	(54.043)	(55.710)	(55.294)	(53.705)	(51.599)	(52.842)
N	2240	5762	8002	4901	4591	9492

Notes: Means reported are for the 2000-2004 estimation sample with standard deviations are in parenthesis. All monetary figures are in year 2000 prices. Contractual gross wage rate and both measures of hours are for combined primary and secondary jobs. See Table A1 for a detailed description of each variable.

Table 2: Labor Supply Function: Difference-in-Difference Approach, Males

	OLS	FE	IV-FE	RF-FE
Log (contractual gross wage rate)	-0.088***	-0.079***	-0.222***	…
	(0.006)	(0.009)	(0.086)	
Age	0.013***	0.040**	0.060**	0.029
	(0.003)	(0.020)	(0.025)	(0.020)
Age squared	-0.018***	-0.027**	-0.042***	-0.018
	(0.004)	(0.012)	(0.016)	(0.013)
Years of schooling	0.001	-0.005	-0.003	-0.007
	(0.002)	(0.004)	(0.005)	(0.004)
Married	0.037***	-0.012	-0.003	-0.016
	(0.012)	(0.024)	(0.026)	(0.025)
1 or 2 children in HH	0.025**	0.032**	0.032**	0.031**
	(0.010)	(0.014)	(0.014)	(0.014)
3 or more children in HH	-0.021	-0.057	-0.070	-0.050
	(0.032)	(0.067)	(0.075)	(0.067)
Child <7 years old in HH	-0.017	0.000	-0.006	0.003
	(0.011)	(0.015)	(0.016)	(0.015)
Log (non-labor income)	0.008**	-0.003	-0.001	-0.004
	(0.004)	(0.005)	(0.005)	(0.005)
Labor only source of income	0.121***	-0.003	0.022	-0.017
	(0.030)	(0.036)	(0.042)	(0.036)
Year = 2001	-0.017	-0.036	-0.018	-0.039
	(0.015)	(0.024)	(0.028)	(0.024)
Year = 2002	0.004	-0.035	0.004	-0.049
	(0.016)	(0.039)	(0.048)	(0.039)
Year = 2003	0.012	-0.039	0.010	-0.058
	(0.016)	(0.055)	(0.067)	(0.054)
Year = 2004	0.013	-0.058	0.000	-0.084
	(0.016)	(0.073)	(0.087)	(0.070)
Treat	0.035*	…	…	…
	(0.019)			
Treat × post reform dummy	0.073***	0.052***	0.053***	0.051***
	(0.018)	(0.019)	(0.019)	(0.019)

Notes: N=8002. We report Arellano (1987) robust standard errors in parentheses to control for arbitrary serial correlation; *** significant at 1% level, ** significant at 5% level and * significant at 10% level. Dependent variable is the log of actual hours worked at primary and secondary jobs in last 30 days. Wages and income are in constant 2000 prices. All specifications are estimated on the full sample (i.e, all individuals with positive labor earnings) for years 2000–2004. *Treat* is the treatment dummy (=1 if treated) defined using post reform actual earnings. "No children" is the omitted category for number of children in HH. HH denotes household. FE is the estimator with individual fixed effects. The reduced form specification (RF-FE) includes the following demand shifters: dummies for 9 categories of firm size (# of employees), the log of difference in regional employment, the log of difference in regional value added per capita, and the local employment share of foreign firms. The demand shifters are used as instruments in the IV specification. The first stage F-statistics (p-value) is 3.016 (0.001) and Hansen's J statistics (p-value) is 11.376 (0.329). R^2=0.056 in OLS and 0.029 in FE (within).

Table 3: Labor Supply Function: DID Approach, Females

	OLS	FE	IV-FE	RF-FE
Log (contractual gross wage rate)	-0.109***	-0.105***	-0.072	…
	(0.008)	(0.011)	(0.213)	
Age	0.012***	0.016	0.013	0.006
	(0.004)	(0.025)	(0.032)	(0.025)
Age squared	-0.014***	-0.017	-0.013	-0.003
	(0.005)	(0.014)	(0.033)	(0.014)
Years of schooling	-0.001	-0.004	-0.005	-0.006
	(0.002)	(0.005)	(0.006)	(0.005)
Married	-0.050***	-0.029	-0.029	-0.029
	(0.009)	(0.019)	(0.019)	(0.019)
1 or 2 children in HH	-0.006	-0.004	-0.005	-0.008
	(0.010)	(0.014)	(0.017)	(0.014)
3 or more children in HH	-0.104**	0.018	0.022	0.030
	(0.048)	(0.091)	(0.094)	(0.093)
Child <7 years old in HH	-0.045***	-0.040**	-0.039**	-0.037**
	(0.012)	(0.016)	(0.018)	(0.016)
Log (non-labor income)	0.003	-0.002	-0.002	-0.003
	(0.004)	(0.005)	(0.005)	(0.005)
Labor only source of income	0.038	-0.035	-0.040	-0.050
	(0.034)	(0.038)	(0.047)	(0.039)
Year = 2001	0.001	0.007	0.001	-0.011
	(0.014)	(0.025)	(0.046)	(0.026)
Year = 2002	0.027*	0.042	0.027	-0.005
	(0.014)	(0.045)	(0.109)	(0.046)
Year = 2003	0.032**	0.048	0.028	-0.013
	(0.014)	(0.066)	(0.141)	(0.067)
Year = 2004	0.043***	0.056	0.032	-0.021
	(0.014)	(0.086)	(0.178)	(0.087)
Treat	0.138***	…	…	…
	(0.018)			
Treat × post reform dummy	0.024	0.002	-0.001	-0.007
	(0.018)	(0.018)	(0.026)	(0.018)

Notes: N=9492. We report Arellano (1987) robust standard errors in parentheses to control for arbitrary serial correlation; *** significant at 1% level, ** significant at 5% level and * significant at 10% level. Dependent variable is the log of actual hours worked at primary and secondary jobs in last 30 days. Wages and income are in constant 2000 prices. All specifications are estimated on the full sample (i.e, all individuals with positive labor earnings) for years 2000–2004. *Treat* is the treatment dummy (=1 if treated) defined using post reform actual earnings. "No children" is the omitted category for number of children in HH. HH denotes household. FE is the estimator with individual fixed effects. The reduced form specification (RF-FE) includes the following demand shifters: dummies for 9 categories of firm size (# of employees), the log of difference in regional employment, the log of difference in regional value added per capita, and the local employment share of foreign firms. The demand shifters are used as instruments in the IV specification. The first stage F-stat (p-value) is 3.452(0.016) and Hansen's J statistics (p-value) is 0.210(0.9). R^2=0.065 in OLS and 0.027 in FE (within).

Table 4: Labor Supply Response to Flat Tax: DID Approach, IV-FE

	Males		Females	
	Actual hours	**Usual hours**	**Actual hours**	**Usual hours**
Primary job	0.045**	0.031**[a]	-0.014	0.007
	(0.019)	(0.012)	(0.025)	(0.014)
	[7973]	[8808]	[9452]	[10516]
Primary and secondary job	0.053***	0.041***[a]	-0.001	0.020
	(0.019)	(0.013)	(0.026)	(0.015)
	[8002]	[8817]	[9492]	[10535]
All jobs	0.067***	0.048***[a]	0.002	0.025
	(0.022)	(0.014)	(0.029)	(0.015)
	[8007]	[8812]	[9496]	[10533]
Including 1998	0.050***[a]	0.046***[a]	-0.002	0.022
	(0.016)	(0.012)	(0.026)	(0.014)
	[10493]	[11540]	[12472]	[13867]
Including 2005	0.050***	0.044***[a]	0.007	0.010
	(0.019)	(0.013)	(0.024)	(0.013)
	[9529]	[10448]	[11277]	[12495]
Adding *district* × *year* interactions	0.061***	0.056***	-0.011	0.002
	(0.022)	(0.016)	(0.024)	(0.014)
	[8002]	[8817]	[9492]	[10535]
Restricted sample	0.046**	0.040***[a]	0.008	0.019
	(0.020)	(0.013)	(0.027)	(0.015)
	[7157]	[7884]	[9138]	[10134]
Contractual earnings	0.056***	0.043***	-0.002	0.014
	(0.019)	(0.013)	(0.026)	(0.015)
	[8195]	[9063]	[9659]	[10786]

Notes: We report Arellano (1987) robust standard errors in parentheses to control for arbitrary serial correlation; *** significant at 1% level, ** significant at 5% level and * significant at 10% level. Number of observations is reported in square brackets. All specifications include individual fixed effects and the same set of controls as in Tables 2 and 3. Estimation is for the years 2000–2004 unless otherwise stated. The instruments are the same as in Tables 2 and 3. Dependent variables are indicated at the top of each column (in logs). Only the estimated treatment effect is reported. The treatment and control groups are defined using actual earnings for the full sample unless otherwise stated. Full sample is defined in notes to Table 2. The restricted sample includes individuals earning between 4,800 and 100,000 rubles per year. District refers to primary sampling units (psu). Models are estimated using hours worked at primary and secondary job unless otherwise stated. The first stage F-test is statistically significant for excluded instruments in all specifications. Except for estimates indicated by *a*, the null hypothesis that the instruments are exogenous is not rejected at 10% level. The observed variation in the sample size is due to different samples and missing values in hours at some jobs.

Table 5: Labor Supply Response to Flat Tax: Weighted DID Approach, IV-FE

	Males		Females	
	Actual hours	**Usual hours**	**Actual hours**	**Usual hours**
Primary job	0.038	0.033**	-0.022	-0.016
	(0.026)	(0.017)	(0.028)	(0.015)
	[7906]	[8733]	[9415]	[10459]
Primary and secondary job	0.047*	0.037*	-0.004	0.002
	(0.028)	(0.020)	(0.028)	(0.016)
	[7935]	[8743]	[9455]	[10479]
All jobs	0.054*	0.041**	-0.001	0.002
	(0.029)	(0.020)	(0.029)	(0.016)
	[7940]	[8738]	[9459]	[10477]
Including 1998	0.041*	0.041**[a]	0.002	0.007
	(0.022)	(0.018)	(0.022)	(0.013)
	[10412]	[11448]	[12427]	[13801]
Including 2005	0.040	0.037*	0.001	-0.005
	(0.027)	(0.020)	(0.026)	(0.015)
	[9456]	[10368]	[11239]	[12439]
Adding *district* × *year* interactions	0.062**	0.046**	0.013	-0.010
	(0.027)	(0.020)	(0.032)	(0.020)
	[7935]	[8743]	[9455]	[10479]
Restricted sample	0.047*	0.037*	-0.004	0.002
	(0.028)	(0.020)	(0.028)	(0.016)
	[7154]	[7880]	[9137]	[10133]
Contractual earnings	0.053**	0.040**	0.009	-0.006
	(0.026)	(0.018)	(0.025)	(0.016)
	[8165]	[9025]	[9641]	[10755]

Notes: We report Arellano (1987) robust standard errors in parentheses to control for arbitrary serial correlation; *** significant at 1% level, ** significant at 5% level and * significant at 10% level. Number of observations is reported in square brackets. All specifications include individual fixed effects and the same set of controls as in Tables 2 and 3. Estimation is for the years 2000 – 2004 unless otherwise stated. The instruments are the same as in Tables 2 and 3. Dependent variables are indicated at the top of each column (in logs). Only the estimated treatment effect is reported. The treatment and control groups are defined using actual earnings for the full sample unless otherwise stated. The restricted sample includes individuals earning between 4,800 and 100,000 rubles per year. District refers to primary sampling units (psu). Models are estimated using hours worked at primary and secondary job unless otherwise stated. We reject the exogeneity of the instruments in the male specifications when 2005 is included using usual hours. The first stage F-test is statistically significant for excluded instruments in all specifications. Except for estimates indicated by *a*, the null hypothesis that the instruments are exogenous is not rejected at 10% level. A fewer number of observations in Table 5 relative to Table 4 is due to a zero weight given to the observations that are far away from the threshold.

Table 6: DID Estimates of Transitional Probabilities for Extensive Margins, Reduced Form

	Males		Females	
	(1)	(2)	(3)	(4)
	Full sample			
A: Worked at least one hour last month at primary job	0.090**	0.111***	0.126***	0.136***
	(0.037)	(0.035)	(0.027)	(0.028)
	[2448]	[2448]	[4128]	[4128]
B: Hold multiple jobs	-0.008	-0.002	0.018**	0.022***
	(0.011)	(0.010)	(0.008)	(0.008)
	[5907]	[6078]	[7306]	[7492]
	Restricted sample			
A: Worked at least one hour last month at primary job	0.084**	0.108***	0.125***	0.136***
	(0.037)	(0.035)	(0.027)	(0.028)
	[2339]	[2339]	[4091]	[4086]
B: Hold multiple jobs	-0.010	-0.002	0.020**	0.023***
	(0.011)	(0.011)	(0.009)	(0.009)
	[5402]	[5656]	[7073]	[7301]

Notes: We report Arellano (1987) robust standard errors in parentheses to control for arbitrary serial correlation; *** significant at 1% level, ** significant at 5% level and * significant at 10% level. Number of observations is reported in square brackets. Dependent variables are indicated in the first column (all dummies). Reported is the effect of the reform on the transitional probability for each margin. All specifications include individual fixed effects and the same set of variables as in Table A3. Full sample is defined in notes to Table 2. The restricted sample includes individuals earning between 4,800 and 100,000 rubles per year. Estimation in *A* is for the years 2001–2004 conditional on non-participation in 2000. Estimation in *B* is for the years 2001–2004 conditional on having a single job in 2000. The treatment and control groups are defined using actual earnings (models 1 and 3) and contractual earnings (models 2 and 4). Labor earnings for working-age non-participants in *A* are predicted using gender, age, age squared, years of schooling, marital status, number of children, non-labor household income, regional labor demand shifters, and 38 dummies for municipalities.

Table A1: Variable Description

Variable Name	Definition
Individual Labor Supply (actual hours h^a)	
At primary job	Hours worked last month at primary job; unusually high hours are top coded at 480 hours per month (16 hours per day*30 days)
At primary and secondary job	Hours worked last month at primary job + hours worked last month at secondary job
At all jobs	Hours worked last month at primary job + hours worked last month at secondary job + hours spent last month on individual economic activities (activities for which an individual is paid, such as sewing a dress, assisting with repairs, selling goods in a market or on the street, etc.)
Individual Labor Supply (usual hours h^u)	
At primary job	4 times usual hours in a typical week at primary job
At primary and secondary job	4 times usual hours in a typical week at primary job + 4 times usual hours in a typical week at secondary job
At all jobs	4 times usual hours in a typical week at primary job + 4 times usual hours in a typical week at secondary job + hours spent last month on regular individual economic activities.
Individual Labor Supply (extensive margins)	
Worked one hour	= 1 if worked at least one hour at primary job last month (defined using actual hours)
Hold multiple jobs	= 1 if individual had more than one job last month
Individual Net Labor Earnings	
Actual net earnings	Money received last month from primary job + money received last month from secondary job
Actual net wage rate (w^a)	Actual net earnings divided by the corresponding *actual hours.*
Contractual net earnings	Monthly average labor earnings from primary job + money received last month from secondary job
Contractual net wage rate (w^c)	Contractual net earnings divided by the respective *usual hours*.
Individual Gross Declared Labor Earnings	
Contractual gross earnings	Gross declared earnings, Y^G, are computed by inverting the tax function, T, for each year t based on reported net earnings, Y^N, and standard deductions, D: $Y_{it}^N = Y_{it}^G - T_t(Y_{it}^G - D_{it})$. Similar to IKK (2005), we assume that earnings are received evenly throughout the year when inverting the tax function.
Contractual gross wage rate	Contractual gross earnings divided by the respective *usual hours*.
Non-Labor Income	
Non-labor family income	Disposable household income minus own contractual labor earnings, where disposable household income is the sum of contractual earnings from all sources across all individual respondents within the household + imputed contractual labor earnings for working-age non-respondents within the household + net private transfers + financial income received last month + public transfers.

Other Variables	
Household composition	Vector of household composition variables: 3 categories for the number of children 16 years old and younger (0, 1-2, and 3+) and a dummy for individuals with a child 7 years or younger in HH.
Married	=1 if married
Age	Continuous age variable.
Schooling	Years of schooling
Change in regional employment	Annual difference in the log of regional employment
Change in regional value added per capita	Annual difference in the log of regional value added per capita
Local share of foreign firms	Share of foreign firms in local employment (at the psu level)
Size of firm	9 categories for firm size

Table A2: Summary Statistics for Demand Shifters

Variables	Mean	Std. Deviation
Firm size categories		
1-5 employees	0.044	0.205
6-20 employees	0.123	0.329
21-50 employees	0.134	0.341
51-200 employees	0.183	0.387
201-500 employees	0.094	0.292
501-2000 employees	0.102	0.303
> 2000 employees	0.053	0.225
Self-employed with no employees	0.061	0.240
Firm size missing	0.205	0.404
Change in regional employment	0.006	0.015
Change in regional value added per capita	0.247	0.091
Local share of foreign firms	0.043	0.046

Notes: N=18686. Means reported are for the period 2000-2004. The sample includes males and females as well as treated and controlled individuals. Variables are defined in Table A1.

Table A3: Labor Supply Response to Flat Tax: DID Approach, Reduced Form

	Males		Females	
	Actual Hours	**Usual hours**	**Actual hours**	**Usual hours**
Primary job	0.044**	0.030**	-0.021	-0.013
	(0.018)	(0.013)	(0.018)	(0.011)
	[7973]	[8808]	[9452]	[10516]
Primary and secondary job	0.051***	0.039***	-0.007	-0.003
	(0.019)	(0.014)	(0.018)	(0.011)
	[8002]	[8817]	[9492]	[10535]
All jobs	0.050***	0.038***	-0.007	-0.001
	(0.019)	(0.014)	(0.018)	(0.011)
	[8007]	[8812]	[9496]	[10533]
Including 1998	0.046***	0.039***	-0.011	-0.008
	(0.016)	(0.011)	(0.015)	(0.010)
	[10493]	[11540]	[12472]	[13867]
Including 2005	0.054***	0.045***	-0.003	-0.006
	(0.019)	(0.014)	(0.017)	(0.011)
	[9529]	[10448]	[11277]	[12495]
Adding *district* × *year* interactions	0.049**	0.046***	-0.006	0.003
	(0.021)	(0.016)	(0.019)	(0.012)
	[8002]	[8817]	[9492]	[10535]
Restricted sample	0.048**	0.039***	-0.004	-0.004
	(0.019)	(0.014)	(0.018)	(0.011)
	[7157]	[7884]	[9138]	[10134]
Contractual earnings	0.042**	0.035***	-0.002	-0.010
	(0.019)	(0.013)	(0.018)	(0.011)
	[8195]	[9063]	[9659]	[10786]

Notes: We report Arellano (1987) robust standard errors in parentheses to control for arbitrary serial correlation; *** significant at 1% level, ** significant at 5% level and * significant at 10% level. Number of observations is reported in square brackets. All specifications include individual fixed effects and the same set of controls as in Tables 2 and 3. Estimation is for the years 2000–2004 unless otherwise stated. Dependent variables are indicated at the top of each column (in logs). Only the estimated treatment effect is reported. The treatment and control groups are defined using actual earnings for the full sample unless otherwise stated. Full sample is defined in notes to Table 2. The restricted sample includes individuals earning between 4,800 and 100,000 rubles per year. District refers to primary sampling units (psu). Models are estimated using hours worked at primary and secondary job unless otherwise stated. The observed variation in the sample size is due to different samples and missing values in hours at some jobs. The number of observations is the same as in Table 4.

Table A4: Labor Supply Response at Different Thresholds, DID Approach

	Males		Females	
	Actual hours	**Usual hours**	**Actual hours**	**Usual hours**
		Panel A: Reduced Form		
Upper brackets × D_p	0.047**	0.032**	-0.007	-0.010
	(0.019)	(0.014)	(0.018)	(0.011)
31% top bracket × D_p	0.015	0.030	0.001	0.046**
	(0.028)	(0.020)	(0.043)	(0.020)
	[8002]	[8817]	[9492]	[10535]
		Panel B: Instrumental Variables		
Upper brackets × D_p	0.048**	0.034**	-0.001	0.012
	(0.020)	(0.014)	(0.027)	(0.015)
31% top bracket × D_p	0.025	0.034*	-0.000	0.046*
	(0.029)	(0.018)	(0.043)	(0.026)
	[8002]	[8817]	[9492]	[10535]

Notes: We report Arellano (1987) robust standard errors in parentheses to correct for arbitrary serial correlation; * significant at 10%; ** significant at 5%; *** significant at 1%. Number of observations is reported in square brackets. Dependent variables are indicated at the top of each column (in logs). All specifications include the same set of variables as in Tables 2 and 3, year and individual fixed effects, and are estimated on the full sample with actual earnings used to define treatment and control groups. As with Tables 2 and 3, the demand shifters in the reduced form models are used as instruments in the IV specifications. Reported are the estimated coefficients on the interaction of a post-reform dummy with the two upper brackets and the 31% top bracket. D_p denotes a dummy for the post-reform period.