9-22-2014

Paying for Altruism: The Case of Organ Donation Revisited

Firat Bilgel
Okan University, firat.bilgel@okan.edu.tr

Brian D. Galle
Boston College Law School, brian.galle@law.georgetown.edu

Paying for Altruism: The Case of Organ Donation Revisited

Fırat Bilgel* Brian Galle†

Abstract

Although many commentators have called for increased efforts to incentivize organ donations, theorists and some evidence suggest these efforts will be ineffective or even could perversely crowd out altruistic efforts. Prior papers examining the impact of tax incentives for donations generally report zero or negative coefficients. We argue these studies incorrectly define their tax variables, and rely on difference-in-differences methods despite likely failures of the requisite parallel trends assumption. We therefore aim to identify the causal effect of tax incentive legislation to serve as an organ donor on living related and unrelated kidney donation rates in the U.S states using more precise tax data and allowing for heterogenous and time-variant causal effects. Employing a synthetic control method, we find that the passage of tax incentive legislation increased living unrelated kidney donation rates by about 52 percent in New York relative to a comparable synthetic New York in the absence of legislation. We show that this causal effect is robust to the exclusion of any particular state as well as to the use of a very small number of comparison states.

JEL Classification: I18; K32; C15

Keywords: Living kidney donation, altruism, tax deduction, difference-in-differences, LIML, synthetic control

*Corresponding Author. Faculty of Economics and Administrative Sciences, Okan University Tuzla Campus Akfırat 34959 Istanbul Turkey. Email: firat.bilgel@okan.edu.tr.

†Boston College Law School, 885 Centre Street, Newton, MA 02459 United States. Email: brian.galle@bc.edu.

1 Introduction

Can altruism be bought? Prior literature, dating back to Olson (1971), emphasizes the importance of "selective incentives" or private goods in encouraging individual contributions to group goals. At the same time, some studies, many in the laboratory, suggest that explicit monetary incentives may have minimal or even perverse effects.[1] Titmuss (1970) famously argued that paying for blood donations would undermine social norms of generosity and reduce giving.

Organ donations are a policy area where these theoretical questions are especially pressing. Earlier qualitative research generally confirms that non-monetary selective incentives, such as the development of emotional and associative bonds between donors and donees, can contribute importantly to expanding the donor base.[2] Perhaps inspired by these efforts, a number of U.S. states have enacted tax incentives for living organ donors.

In four previous studies, however, researchers employing difference-in-differences (DiD) methods have consistently failed to find any evidence that these incentives affected kidney donations, which comprise the large majority of demand for and supply of donations. Wellington and Sayre (2011) report that state legislation is not associated with overall living donations. Venkataramani et al. (2012) find no statistically significant contemporaneous or lagged effects of tax policy on donation rates, or differential effects by gender, race or donor relationship. They hypothesize that the statistically indistinguishable effects may stem from low cash value of the tax deduction to defray costs faced by donors, lack of public awareness and a depletion of organ donor pools in the pre-legislation period. Lacetera et al. (2014) find significant effects for bone marrow but not other organs. Boulware et al. (2008) find that state legislation and federal policies are not associated with living related or overall donations. They do find a positive effect on living unrelated kidney donations when pooling together all state legislation, but do not separately analyze monetary incentives.

While all four studies largely reach a consensus, they share a number of methodological and data issues in the identification of causal effects of the law that leave their conclusions open to question.

Pooling cross-section and time-series data may cloud any potentially significant effect one might have observed if these effects were analyzed by a state-by-state pure time-series analysis. On the other hand, a pure time-series analysis of the law and kidney donations would be contaminated by structural shocks.

Compositional differences and non-parallel trends may pose threats to the validity of DiD estimation. That is, the distribution of the law and the covariates that are thought to affect kidney donation rates may not be similar for the pre- and post-legislation periods or the treatment (i.e. states that passed the legislation) and control states (i.e. no legislation) may not have experienced the same trends in the absence of the law conditional on covariates. As Boulware et al. (2008) show graphically, the trend lines for enacting and non-enacting states cross prior to the enactment period (though Boulware et al. (2008) do not note the econometric significance of that fact). Further, there are good reasons to suspect that tax incentives tend to be enacted near the peak of public attention to the organ donation crisis, implying the likelihood of regression to the mean in enacting but not necessarily non-enacting states.

The prior studies also face a set of other identification challenges. Each treats enactment of tax legislation as a binary variable, when in fact the real dollar value of the incentives states offer ranges from a few hundred to ten thousand dollars. In addition to failing to account for this variation, prior papers overlooked two states that enacted credits, not deductions. Next, 70 percent of all U.S states have passed either a paid leave of absence and/or a tax deduction/credit legislation between the 1990s and 2010. For a state that subsequently

[1] Gneezy et al. (2011) offer a review.
[2] See Healy (2010) for a review.

passed both of these legislations, one may not be able to isolate the causal effect of one from the other.

Lastly, the enactment of tax incentive legislation may not be exogenous to kidney donations. In this case, one should resort to an instrumental variable procedure where another variable should be found such that it moves around the covariate of interest (i.e. the law) in a way that can plausibly be viewed as random. We argue that finding such an exogenous source of variation - a plausible instrument for the enactment and value of each state's tax incentive - is very difficult.

We therefore propose another strategy that is robust to the above-mentioned problems provided that its identifying assumptions are met. We employ the synthetic control method (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2014) where we create a synthetic control group that replicates the-pre legislation living kidney donation rates of the states that enacted a tax incentive legislation by using a convex combination of other states that have not enacted any legislation. For the sake of completeness and for purposes of cross-validation we also perform (1) a DiD approach *à la* Bertrand et al. (2004) where the tax incentive legislation is captured by a binary variable, and (2) a limited information maximum likelihood (LIML) estimation where the real dollar value of the legislation is measured accurately and captured an endogenous treatment.

We find no statistically significant causal effect of tax incentive policies on related or unrelated kidney donation rates via a DiD estimation or an instrumental variable procedure which accounts for the endogenous nature of the law. If anything, we find some weak evidence that enactment of tax incentives diminishes related donations. While in theory this could represent crowding-out of altruistic behavior by money incentives, we cannot rule out the possibility that instead we are observing mean regression.

When we account for this problem through a synthetic control method that allows for unobservable state heterogeneity to vary over time, we find that the passage of tax incentive legislation increased living unrelated kidney donation rates by about 52 percent in New York relative to a comparable synthetic New York in the absence of legislation. We show that this causal effect is robust to the exclusion of any particular state as well as to the use of a very small number of comparison states. It is possible that New York is unique, but our methodology does not allow us to measure accurately effects in other states, so that we can neither rule out nor confirm the efficacy of tax incentives elsewhere.

In short, our advances over prior literature are primarily to refine the methodological approach and the legal accuracy of the data. Unlike earlier papers, our DiD techniques account for several possible dimensions of endogeneity in the data, and accurately account for large variations in the value of the incentives offered. We also show reasons to believe DiD estimates will tend to produce spurious negative correlations, and when using methods robust to dropping the DiD assumptions find evidence that tax incentive legislation may increase donations.

Section 2 discusses the background on state legislation, section 3 introduces our empirical strategy where we respectively define tax incentive legislation as a binary treatment, a non-binary endogenous treatment and finally allow for heterogenous causal effects of the law, section 4 discusses the findings of the analysis with respect to methodological differences and identification strategies and section 5 concludes.

2 Background and Prior Literature

2.1 Statutory Background

In 1984, the U.S. Congress enacted the National Organ Transplantation Act (NOTA)[3]. NOTA established a network of sub-national organ procurement organizations (OPO), each with jurisdiction spanning about

[3]See Satel and Hippen (2007) for more details on NOTA.

one state on average. These OPOs have primary responsibility for soliciting organ donations and matching patients in need with eligible donors.

NOTA also prohibited trade in organs, but at the time of its initial enactment it was unclear whether this prohibition extended to bar reimbursement of donor expenses. During the 1990s, state government interest in encouraging organ donation swelled, but NOTA was seen as a possible obstacle to reform. For example, in 2000 the Kansas Attorney General opined that NOTA prohibited a proposed law allowing tax deductions for expenses associated with organ donation (Calandrillo, 2004). Congress responded in 2000 by amending NOTA to clarify that reimbursements for lost wages, travel, and medical expenses associated with donating would not contravene the Act.

Shortly thereafter, states began enacting moderate financial incentives for donations. Wisconsin was the first, in 2000 granting its own employee donors up to 30 days of paid time off. Thirty other states followed suit, most between 2001 and 2005.

Wisconsin was also the first state to adopt a reimbursement law, enacting a $10,000 tax deduction for NOTA-permitted expenses in 2004. Fourteen states followed, eight of them in 2005, and most similarly providing for a deduction against taxable income to the extent of covered expenses. The laws varied slightly in some details, such as whether undeducted expenses could be carried forward to another tax year; Idaho and Virginia provided for only a maximum $5,000 benefit. More significantly, Idaho, Louisiana and Utah allowed for a credit, rather than a deduction, greatly increasing the actual dollar value of the incentive.

New York is of particular interest for methodological reasons we explain in more detail in Part 3.4. New York adopted paid leave of absence for state employees in 2001. Bilgel (2014) finds no effect of that enactment on donations. In February of 2004, New York passed a resolution awarding a medal of recognition for living and deceased organ donors past and future. It then enacted a $10,000 deduction for donations in 2006.

2.2 Prior Literature

Boulware et al. (2008) were the first to examine whether these state policies were effective in encouraging donations. Like others that followed, and us here, they focused on kidney donations. Kidneys are by far the most frequently donated and demanded organ: approximately two-thirds of all organ donation waiting lists are made up of patients awaiting kidneys. Boulware et al. (2008) pool together all the different state interventions, and find no significant effects of enactment on cadaveric or living related donations. They do, though, find that statistically significant impact of enactment, of about 26%, for living related donations. They also report that growth in donation rates overall dropped sharply after a state's first enactment of any legislation, and that trends in end-stage renal disease between 1988 and 2005 appear to predict state enactment.

Later studies aimed specifically at the effects of tax legislation found no significant effects. Wellington and Sayre (2011) find no significant impact of tax incentives. Venkataramani et al. (2012) further examine lags of the legislation and break out effects by demographic sub-groups, and still find no effects. Lacetera et al. (2014) expand their analysis to include other organs as well as bone marrow, but find no effects for kidney or other organ donations. All three studies rely on DiD methods and OLS. Lacetera et al. (2014) control for lags of factors that might lead to enactment of legislation, such as lagged ESRD rates, but do not actually use instruments for enactment of the legislation.

For the most part these studies treat all tax incentives uniformly. Venkataramani et al. (2012) and Lacetera et al. (2014) note that Idaho offers a tax credit instead of a tax deduction, and report that their results are robust to omitting Idaho. Wellington and Sayre (2011) provide an alternative specification in

which they estimate the actual dollar value of a deduction by multiplying the maximum deduction amount by the top marginal tax rate in the state. They do not identify any states that grant credits. As we noted in section 2.1, in fact Louisiana and Utah also have credits. We also doubt that the Wellington and Sayre (2011) approach closely approximates the dollar value of the state incentive; few donors likely are in the top state bracket, not all donors itemize (and hence not all donors are even eligible to claim a deduction), and federal deductibility of state taxes reduces the net value of the state incentive.

In addition to some question about the measurement of the tax variables, these earlier studies face potential endogeneity issues, as well as with the choice of the DiD design. Since the urgency of obtaining living donors is reduced when there are more plentiful cadaveric donors, and since controlling for cadaveric donors may capture underlying trends in efforts to spur donations generally, it is sensible, as these papers do, to control for cadaveric donations. But the urgency of obtaining cadaveric donations is also reduced when living donations increase, suggesting the two variables may be endogenous to each other.

Further, the enactment of the tax incentives themselves may be correlated either with underlying levels or trends in donations, or both enactment and donation levels may be driven by an unobservable such as the organizational savvy of needy donees. Along these lines, Satel and Hippen (2007) offer an anecdote in which the father of one child in need of a liver leveraged connections with members of Congress to obtain publicity for his daughter's plight.

Finally, we suspect that DiD approaches will be biased and inconsistent because of a failure of the common trends assumption. States that will ultimately enact tax incentive legislation have higher average levels of donations even years before enactment of the legislation. This raises the danger that enacting states may regress to the mean around the time of enactment, violating the DiD assumption that state trends would have been parallel if not for the treatment. Further supporting this inference, pre-enactment trends are not fully parallel, though certainly similar. In some periods the two trend lines are moving in opposite directions. If states are pooled by whether they ever enact any form of organ donor legislation, as in Boulware et al. (2008), then trend lines actually cross prior to the first enactment date.

3 Empirical Strategy

3.1 Data and Sample

We use state-level panel data for the period 1988-2012. State data on the number of related and unrelated living adult kidney donors, kidney waiting list additions (candidates) and the number of transplant centers are retrieved from the Organ Procurement and Transplantation Network (OPTN)[4]. Related donation is composed of donations by blood related child, full sibling, half sibling, identical twin, other relative, parent, spouse and life partner. Unrelated donation are composed of non-biological anonymous donations and other unrelated directed donations only[5].

Total state population and the population over the age of 18 are obtained from the US Census Bureau[6]. The number of living related and unrelated kidney donors, the number of deceased kidney donations and the number of transplant centers are measured in per million adult population (pmap) rates and the kidney

[4] http://optn.transplant.hrsa.gov/

[5] Pairwise kidney exchanges (PKEs) have been excluded from unrelated donations although they are coded as such by OPTN. The reason for excluding PKEs is that the decision to donate by a relative to a biologically unrelated person under pairwise kidney exchange is conditioned upon his/her recipient receiving a kidney from another biologically unrelated person. Thus the motivation under pairwise exchanges is not based upon the same reasons to donate under typical unrelated donations.

[6] http://www.census.gov/popest/states.html

waiting list additions are measured in per million population (pmp) rates. The prevalence of ESRD pmp is retrieved from the US Renal Data System (USRDS) 2010 Annual Report[7]. The real GDP per capita (in 2005 US dollars) and other state fiscal data are obtained from the US Department of Commerce, Bureau of Economic Analysis and the Census Bureau[8]. The number of traffic fatalities are retrieved from the National Highway Traffic Safety Administration, Fatality Analysis Reporting System (FARS)[9]. The number of cerebrovascular deaths are retrieved from CDC-WONDER[10]. Both variables are expressed in pmp rates.

3.1.1 Construction of the Tax Variables

We hand-collect the details of each state's tax incentives for organ donation. We double-check our coding against compilations by the National Kidney Foundation[11], TransplantLiving[12], National Conference of State Legislatures[13], Boulware et al. (2008) and Lacetera et al. (2014).

To generate the true dollar value of the incentives, we employed two alternative methods. Our results are robust to either method. In the simpler approach, we rely primarily on the average state and federal marginal rates facing all taxpayers in each state-year, as calculated by NBER (2013). In the second approach, reflected in our reported results, we generate the average state and federal marginal rate for an itemizing taxpayer in each state-year cell using NBER's Taxsim calculator. As the inputs for the calculator, we take the weighted mean income, household size, and claimed deductions from state-year aggregates reported by the IRS Statistics of Income. These data are stratified by income band, and we weight the statistics for each band by population and share of itemizers to arrive at an overall state-year weighted average. For states with credits, we omit weighting by share of itemizers, to reflect the fact that credits can be claimed by non-itemizers.

Both state and federal marginal rates affect the final value of the organ donor incentive. State income taxes are deductible from federal taxes, and in some states federal taxes are deductible from the state income tax base. State-level deductions thus reduce state tax but also thereby increase federal tax.[14] In the simpler case where state taxes are deductible federally but not vice-versa, the net dollar value of an incentive, I, is therefore $I\tau_s - (I\tau_s\tau_f) = (1-\tau_f)(I\tau_s)$ where τ indicates the relevant tax rate. Estimating the value of a credit is more complex, because not all credit claimants need be federal itemizers. Since we cannot observe how many credit claimants itemize, we assume as an approximation that they itemize at the average state rate, and treat the expected federal cost of the credit as the product of the credit's value and the odds of itemizing. That is, we multiply the federal cost of state credits by the share of federal taxpayers claiming a deduction for state taxes paid in each state-year cell, and subtract the resulting amount from the value of the credit for each cell.

[7] http://www.usrds.org/reference.htm
[8] http://www.bea.gov/regional/gsp/
[9] www.-fars.nhtsa.dot.gov/States/StatesFatalitiesFatalityRates.aspx.
[10] http://wonder.cdc.gov/controller/datarequest/D72;jsessionid=2D458CFDCFCF17D2F52A54B2A78D71FF
[11] http://www.kidney.org/transplantation/LivingDonors/pdf/LDTaxDed_Leave.pdf
[12] http://www.transplantliving.org/livingdonation/financialaspects/legislation.aspx
[13] http://www.ncsl.org/default.aspx?tabid=13383
[14] It is theoretically possible that some taxpayers will choose to itemize at the state level, due to the availability of the organ-donation incentive, but will not itemize federally. If so, our coefficients for the effect of the tax incentive may be slightly attenuated, as we will be somewhat underestimating the value of the incentive.

3.2 Tax Incentive Legislation as a Binary Exogenous Treatment

3.2.1 Specification

We first investigate whether allowing for possible endogeneity in cadaveric donation rates produces results different than prior literature, and for cross-check and verification purposes include baseline results treating cadaveric donations as exogenous. In these specifications, we treat the tax incentive legislation as a binary treatment and specify the following model in the panel data context:

$$y_{it} = \mu_i + \delta_t + \alpha_i \tau_{it} + \gamma D_{it} + X_{it}^{'} \beta + \varepsilon_{it} \tag{1}$$

where y_{it} denote the living kidney donation rates pmap, i denotes the state $i = 1, \ldots, N$, t denotes time $t = 1, \ldots, T$, μ_i, δ_t and α_i respectively are the state and year fixed effects and state linear trend effects, D is a binary treatment dummy that takes the value of 1 if the tax incentive legislation is in effect in state i at time t and 0 otherwise, β is a $K \times 1$ vector where K is the number time-variant regressors with possibly endogenous elements and ε_{it} is the stochastic disturbance term that comprises idiosyncratic/transitory shocks, measurement errors in y_{it} and aggregation errors. X_{it} consists of the prevalence of ESRD pmp, the real per capita GDP, the number of transplant centers pmap, the kidney waiting list additions pmp and the cadaveric donation rate pmap. For the moment, we assume that the legislation is exogenous to kidney donations but the cadaveric donation rate may not be exogenous.

To account for the possibility that cadaveric donations may be endogenous to living donations, a source of exogenous variation should be found such that it might plausibly be viewed as randomly moving around cadaveric donations. The likelihood of becoming a deceased donor (in the medical sense) is greater for individuals who have been exposed to situations in which irreversible brain injury resulting in brain death is more likely. Consequently, given medical compatibility, victims of motor vehicle accidents and cerebrovascular diseases are suitable cadaveric donor candidates (Bilgel, 2013). There is no evident relation between these two variables and living donations, except through their impact on cadaveric donations. Therefore, the cerebrovascular death rates and the motor vehicle fatalities are identified as sources of exogenous variation that are conjectured to be correlated with living related or unrelated kidney donation rates through and only through cadaveric donations and that are robust candidates to be treated as instruments.

Bertrand et al. (2004) show that with serially correlated outcome variable, the DiD estimator leads to inconsistent standard errors, serious overestimation of t-statistics and significance levels. The simplest solution to this problem is to ignore the time-series information to compute the standard errors by averaging the data before and after the passage of the legislation and run equation (1) in a panel of length two. However, when the laws are staggered over time as various states passed tax incentive legislation at various points in time, we regress y_{it} on state and year fixed effects, linear trends and covariates and then divide the residuals of the states that passed the legislation into two groups and average them for the pre- and the post-legislation period. We then regress these residuals on the treatment dummy in a panel of length two and adjust the standard errors for small sample.

3.2.2 DiD Results

Table 1 shows the estimation results for living related (columns 1 through 3) and unrelated (columns 4 through 6) kidney donation rates with an exhaustive set of diagnostics on endogeneity, instrument relevance, weak identification, instrument validity and linearity. When the excluded instruments are only weakly correlated with the endogenous variable, the instrumental variable estimates will be biased in the same direction as the

OLS and the significance tests will have incorrect size and confidence interval. Therefore, we further report weak identification-robust inference test results.

In order to check for the endogeneity of cadaveric donations, we first perform a DiD estimation via LIML where we instrument cadaveric donation rates with cerebrovascular deaths, motor vehicle fatalities and their squares (columns 1 and 4). While the diagnostic test results indicate that the instruments are relevant (correlated with cadaveric donations), clean (uncorrelated with errors) and correctly excluded, they also indicate that the null hypothesis that cadaveric donations may actually be treated as exogenous cannot be rejected at conventional test levels. We next perform a LIML estimation and disregard the exogeneity of cadaveric donations (columns 2 and 5). As methods with asymptotic foundations tend to perform poorly in small samples, in columns 2 and 5, we report two-way cluster-robust bootstrapped standard errors *à la* Cameron et al. (2011) (henceforth CGM)[15]. Finally, based on the diagnostics in columns 1 and 4, we treat cadaveric donations as exogenous and perform the residual aggregation method, suggested by Bertrand et al. (2004) to overcome serially correlated outcomes and staggered laws (columns 3 and 6).

In line with Wellington and Sayre (2011), Venkataramani et al. (2012) and Lacetera et al. (2014), the DiD results in table 1, irrespective of the estimation strategy, show that the tax incentive legislation does not exert a causal effect on living related or unrelated kidney donation rates. Albeit statistically indistinguishable from zero, a crowding-out is evident under related donations. However, these results most likely suggest a failure of the DiD design. One of the basic identifying assumptions in a DiD study is that treatment and control states would have continued on parallel trends if not for the treatment. As we noted, treatment states here tended to have higher levels of donations prior to enactment. Regression to the mean is to be expected, and would also tend to produce a negative sign for enactment.

We cannot definitively rule out mean regression and similar possibilities as an explanation for our results. For example, it might be that enacting states were also leaders in private methods for encouraging donations, which if true would tend to produce a spurious negative correlation between enactment and donations. A similar story would be that efforts to reach out to donors and to lobby legislators were conducted hand in hand, and that both peaked with successful enactment of tax incentives. Here, too, we would expect a decline in donations following enactment, which again would violate the but-for-enactment common trends assumption.

A potentially serious caveat of these models is that the law may not be exogenous to kidney donations. If true, our DiD estimations will be biased and inconsistent. Therefore, in the next section, we re-define the tax incentive legislation as a non-binary endogenous treatment and invoke an instrumental variable procedure where we search for an exogenous source of variation in the legislation.

3.3 Tax Incentive Legislation as a non-binary Endogenous Treatment

3.3.1 Specification

We next investigate whether accurately measuring the tax variables may produce different results than prior literature. As described in 3.1.1 above, we estimate the actual mean dollar value of each state's incentive, and substitute it for the binary tax variable employed in 3.2. We also allow for the possibility that incentive statutes are endogenous to donation rates. For example, legislators may respond to perceived need, or communities that are effective in organizing donors may also be effective in organizing political activism.

[15]We use `xtivreg2` (Schaffer, 2010) and `weakiv` (Finlay et al., 2013) commands in STATA, available at: http://ideas.repec.org/c/boc/bocode/s456501.html (xtivreg2), http://ideas.repec.org/c/boc/bocode/s457684.html (weakiv).

Therefore, in this section, we treat tax incentive legislation as a non-binary endogenous treatment and specify the following model:

$$y_{it} = \mu_i + \delta_t + \gamma T_{it} + X_{it}^{'}\beta + \varepsilon_{it} \tag{2}$$

where the treatment, T, is the mean value of state incentive for donations, the remaining variables are defined as before and all elements of X are assumed to be uncorrelated with the error term. The treatment might be correlated with factors that the specification does not already capture.

It is difficult to identify convincing instruments for the existence and value of the tax incentives. Boulware et al. (2008) report that trends in ESRD predict state legislative action. We compute the ESRD trend for each state from 1988 to 2002 (two years before the first tax incentive was enacted). We interact this trend with average federal marginal rates facing itemizers in each state, and include the interaction term, the rate, and the squared rate as instruments for the tax price of donating.

Alternately, we replace the trend interaction term with the share of state Medicaid expenses paid by the federal government (state FMAP rate), its square, and the product of FMAP and average tax rate. Organ transplantation is a long-run cost saver (Matas and Schnitzler, 2004). However, in the presence of vertical fiscal externalities, such as the Medicaid matching grant, states have lower incentives to adopt cost-saving legislation. We expect that FMAP will therefore be negatively correlated with donation incentives. Admittedly, both FMAP and mean federal marginal rate face the difficulty that they are in part determined by state income levels, which may also affect organ donations. We do, however, control in several ways for state income level and economic activity. Testing of our instruments produces first-stage F scores of less than 10, hence we employ LIML, which is robust to weak instruments.

3.3.2 LIML Results

Table 2 displays the estimation results for the natural log of living related (columns 1-4) and unrelated (columns 5-8) kidney donation rates. Columns 1 and 5 show our baseline specifications where the mean value of state incentive for donation is instrumented by the IIT marginal rate, the federal medicaid matching grant to state and their squares. In columns 2-4 and 6-8 we perform a robustness check for our baseline specification where we include health spending, unemployment rate, share of itemizers, state median income, share of black population, total population, state medicaid spending and two dummy variables capturing the enactment of paid leave of absence for public and private employees as additional control variables.

While under unrelated donation rates, the sign of the effect of the mean value of state incentive for donation is conformable to our expectations, the effect is statistically indistinguishable from zero at conventional test levels. On the other hand, taken at face value, we find a marginally statistically significant and negative effect of the enactment of tax laws on related donations in column 4 only. Specifically, we find that enactment reduces related donations by about 4.5 percent, significant only at the 10% level[16]. It is possible this outcome represents a crowding-out of altruistic donations by explicit monetary incentives. Prior literature reports some instances of this kind of crowding, mostly in laboratory settings (Gneezy et al., 2011). As discussed in section 3.3.1, however, the negative coefficients may also be artifacts of the DiD design.

We next attempt to implement a research design that allows unobservable state heterogeneity to vary over time.

[16]We additionally performed OLS estimations whose results are slightly smaller in magnitude, with reductions of between 1.2 and 2.3% depending on specification, and significant at the 5% level. We obtain similar results for the effects of tax incentives only in states that also enacted paid leave and for regressions including interactions between the tax and paid leave variables. These additional estimations are available from the authors upon request.

3.4 Heterogenous Causal Effects

In this secton, we attempt to reveal the causal effect of tax incentive legislation by imputing the missing potential outcome, that is the (counterfactual) outcome that would have prevailed if states had not passed the legislation. For this purpose, we invoke the synthetic control method, developed by Abadie and Gardeazabal (2003) and extended by Abadie et al. (2010) and Abadie et al. (2014). Suppose there are $S+1$ states, indexed by $i = 1, 2, \ldots, S+1$ over T periods, $t = 1, 2, \ldots, T$. Only state $i = 1$ passed the legislation and the remaining S states are the potential control states that did not pass any legislation, called the *donor pool*. There are T_0 number of pre-legislation periods and T_1 number of post-legislation periods so that $T_0 + T_1 = T$. The effect of the law for unit i at time t is given by $\alpha_{it} = Y_{it}^I - Y_{it}^N$ where Y_{it}^I is the living kidney donation rate of state i if the legislation is enacted in $T_0 + 1$ to T and Y_{it}^N is the living kidney donation rate in the absence of law. Since only state $i = 1$ enacted the legislation, we need to estimate $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$. We first estimate Y_{it}^N by the following factor model:

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it} \tag{3}$$

where δ_t is an unknown common factor invariant across units, Z_i is the covariate vector not affected by the law, θ_t is a vector of unknown time-specific parameters, λ_t is a vector of unknown common factors, μ_i is the state-specific unobservable and the error term ε_{it} are the zero-mean transitory shocks. The presence of anticipatory effects are irrelevant in our case, implying that all the elements in Z_i that belong to pre-legislation period are unaffected by the law. Equation (3) allows the effect of unobservable state heterogeneity to vary over time. In the panel DiD estimator, the effect of unobservable heterogeneity, λ_t, is assumed to be fixed over time. Hence, the synthetic control method provides an improvement over the DiD method and deals better with endogeneity caused by the presence of time-varying unobservable confounders.

The method aims to construct the missing counterfactual, Y_{it}^N, from states not affected by the law. Let $W = (w_2, \ldots, w_{S+1})'$ be $(S \times 1)$ vector of weights such that $0 \leq w_j \leq 1$ for $j = 2, 3, \ldots, S+1$ and $\sum_{j=2}^{S+1} w_j = 1$. Define the linear combination of pre-legislation values of living kidney donation rates by $\bar{Y}_j^k = \sum_{m=1}^{T_0} k_m Y_{jm}$. Abadie et al. (2010) shows that if the following conditions hold, then the estimate of the effect of the law for the enacted state, $\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{S+1} w_j^* Y_{jt}$, is an unbiased estimator of α_{1t}:

$$\sum_{j=2}^{S+1} w_j^* Z_j = Z_1 \quad \wedge \quad \sum_{j=2}^{S+1} w_j^* \bar{Y}_j^k = \bar{Y}_1^k \tag{4}$$

where w_j^* is the weight assigned to the j^{th} state that did not enact the law.

Equation (4) can hold exactly only if $\left(\bar{Y}_1^k, Z_1\right)$ belongs to the convex hull of $\left[\left(\bar{Y}_2^k, Z_2\right), \ldots, \left(\bar{Y}_{P+1}^k, Z_{P+1}\right)\right]$. This means it is possible that the pre-legislation living kidney donation rate of some of the legislation-enacted states may not be synthetized accurately using the pre-legislation characteristics of the states that did not enact such legislation.

The vector W^* is chosen to minimize the distance between the vector of pre-legislation characteristics for the exposed state (X_1) and the weighted matrix that contains the pre-legislation characteristics of unexposed states (X_0): $\| X_1 - X_0 W \| = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$ where V is a symmetric and positive semidefinite matrix. This minimization procedure is subject to the constraints that the weight assigned to each unexposed state should lie between zero and one and that the sum of the weights is bounded by one.

3.4.1 Treatment and Control Selection

Until 2012, 18 states in the U.S passed tax incentive legislation and 15 states did not enact any type of legislations. Of the 18 states that passed tax incentive legislation, Idaho was discarded due to lack of data on the number of living kidney donors and 14 states (Arkansas, Georgia, Iowa, Minnesota, Missisippi, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Utah, Virginia, Wisconsin) were discarded because they have already enacted paid leave of absence legislation (i.e. one cannot isolate the effect of paid leave and the effect of tax incentive legislation). The only exception to this case is New York where the paid leave of absence legislation (2001) is shown not to exert a causal effect on living unrelated kidney donation rates (Bilgel, 2014). Therefore the effect of the following tax incentive legislation (2006) in New York can be isolated[17]. This leaves three treatment states (Louisiana, Rhode Island and New York) to be used in the synthetic control. On the other hand, of the 15 states that enacted none of the above laws, 2 states (Montana, Wyoming) were discarded due to lack of data on the number of living kidney donors. This leaves a total of 13 states to be considered in the donor pool. Table 3 displays the assignment of treated and donor states.

The set of characteristics used to synthetize the outcome variables, the living related and unrelated kidney donation rates per million adult population (pmap) are the living related and unrelated kidney donation rates of the control states without any legislation for every year in the pre-legislation period, kidney waiting list additions per million population (pmp), deceased kidney donations pmap, the number of transplant centers pmap, the prevalence of end-stage renal disease (ESRD) pmp, traffic fatalities pmp, cerebrovascular deaths pmp and real GDP per capita for the period 1988-2010. The living related and unrelated kidney donation rates were available for each year under consideration, hence used for every year in the pre-legislation period. The remaining covariates are used when available at least for one year in the pre-legislation period.

In the final specification, we dropped all the covariates except the living kidney donation rates of the unexposed control states for every year in the pre-legislation period because neither did other covariates improve the pre-legislation fit of the trajectory of the living kidney donation rates nor yield accurate donation rate trajectories when all the covariates except the living kidney donation rates of the unexposed control states were used. This implies that the pre-legislation actual donation trajectory is best reproduced by some linear combination of the donation rates of the unexposed states.

3.4.2 Synthetic Control Results

Among the three treated states, Rhode Island is discarded from the analysis because the tax incentive legislation is enacted in 2009, leaving only four years to analyze the causal effect of the legislation in the post-legislation period as our sample ends in 2012. The results of the synthetic control and the subsequent placebo studies for Louisiana are not reported either because the synthetic control method was unable to reproduce the pre-legislation donation rate trajectory. We also omit the results of the synthetic controls and the subsequent placebo studies for related donations because the synthetic control method was unable to successfully reproduce the pre-legislation related donation rate trajectory in any of these three treated states. We therefore report here and in the following sections the results of the analysis for New York and for unrelated donation rates only. The entire analysis for Rhode Island and Louisiana or for New York's related donation rates is available from the authors upon request.

The pre-legislation period for New York is 18 years. With a large number of pre-legislation periods,

[17]Employing the synthetic control method, Bilgel (2014) was unable to confirm the presence or the absence of a causal effect of paid leave of absence legislation in 12 out of 16 states in the U.S. Of the remaining four states, the paid leave of absence legislation was found to be effective only in California.

matching on pre-legislation outcomes allows to control for heterogenous responses to multiple unobserved factors. The intuition is that only states that are alike in both observed and unobserved determinants as well as in the effect of those determinants on kidney donation rates should produce similar trajectories of the kidney donation rates over extended periods of time (Abadie et al., 2014).

Figure 1 plots the trends in living unrelated kidney donation rates for New York and its synthetic counterpart over the period 1988-2012. The synthetic living unrelated kidney donation rate trajectory is constructed by using the convex combination of states in the donor pool that closely resembled New York before the passage of tax incentive legislation[18]. The synthetic unrelated donation rate trajectory almost perfectly reproduces the actual unrelated donation rate trajectory in the pre-legislation period. In the next sseven years following the passage of tax deduction legislation, the synthetic unrelated donation rate takes a dive while the actual unrelated donation rate in New York rises.

The estimate of the impact of tax deduction legislation is given by the difference between the actual and the synthetic unrelated kidney donation rates in the post-legislation period. Our findings suggest that in the post-legislation period, the living unrelated kidney donation rates in New York increased on average by about 52 percent relative to a comparable synthetic New York in the absence of legislation[19].

3.4.3 Inference

In order to ensure that a particular synthetic control estimate reflects the impact of the legislation (i.e. the synthetic controls provide good predictors of the trajectory of living kidney donation rate in the pre-legislation periods), we perform a series of falsification tests known as *in-space placebo* test, in which the legislation period is artificially reassigned to each of the 13 states which did not enact tax deduction legislation and shift the treated state into the donor pool (Abadie et al., 2010). If a particular state enacted the legislation and other states did not, our expectation is that the control states that are subject to the synthetic control method should not be affected by the legislation. Thus we should not observe actual and synthetic living kidney donation rates diverge in the post-legislation period. Our confidence that a sizeable synthetic control estimate reflects the effect of legislation would be severely undermined if similar or larger estimated kidney donation rate gaps are obtained when the legislation is artificially assigned to states which did not enact such legislation (Abadie et al., 2010).

The post-legislation living unrelated kidney donation rate gap in New York is the largest of all placebo gaps and thus the effect of tax deduction legislation on living unrelated kidney donation rates in New York is causal rather than a random effect. In order to assess whether the estimated effect is causal, the synthetic control method is applied to estimate *in-space* placebo kidney donation gaps for every potential control state in order to create a distribution of placebo effects. This distribution enables us to identify the exact significance level of the estimated effect of the legislation. Our confidence that a sizeable synthetic control estimate reflects the effect of the legislation would be severely undermined if the estimated gap fell well inside the distribution of placebo gaps (Abadie et al., 2010). This would imply that our results are driven by randomness rather than causality. In other words, a significant causal effect of the legislation in the treated

[18]We use the `synth` command in STATA, which can be found at http://www.stanford.edu/~jhainm/synthpage.html. All synthetic control estimations are performed using the `nested` option. By default `synth` relies on a constrained quadratic programming routine that finds the best fitting W-weights conditional on the regression-based V-matrix. With `nested` option, the `synth` command embarks on a fully nested optimization procedure that searches among all (diagonal) positive semi-definite V-matrices and sets of W-weights for the best fitting convex combination of the control states. See STATA `synth` help file for more details.

[19]This causal effect is calculated by taking the ratio of the difference between the average unrelated donation rate of New York and the average unrelated donation rate of the synthetic New York to the average unrelated donation rate of the synthetic New York in the post-legislation period.

state requires that the estimated effect should be unusually large relative to the distribution of placebo effects. The estimated effect of the legislation for New York is evaluated by calculating the ratio of post-legislation root mean square prediction error (henceforth RMSPE) to pre-legislation RMSPE that are equal to or greater than the one for New York[20]. This ratio is the p-value that can be interpreted as the probability of obtaining a post/pre-legislation RMSPE that is at least as large as the one obtained for New York when the legislation is artificially and randomly reassigned to a state that did not enact such legislation.

Figure 2 plots the distribution of placebo effects for New York and for every 13 states in the donor pool for unrelated donations. The estimated living unrelated kidney donation rate gap, by far, fell well outside the distribution of placebo gaps. This means that, if a state would have been randomly selected from the sample, the probability of obtaining a post/pre-legislation RMSPE ratio as high as that of New York would be $1/14 = 0.0714$. No control state in the sample achieves an equally high ratio.

3.4.4 Robustness Check

In this section, we perform a robustness check to test the sensitivity of the results to the changes in the synthetic control state weights induced by the exclusion of any particular state from the sample as in Abadie et al. (2014). From table 4, the synthetic New York is constructed by the weighted average of nine states, namely Alabama, Arizona, Florida, Michigan, Nebraska, New Jersey, Nevada, Tennessee and Vermont. We iteratively re-estimate our model to construct a synthetic New York excluding in each iteration one of the states that was assigned a weight in table 4. We aim to assess the extent to which our results are driven by any particular state. Figure 3 displays the results in which the black solid line is the actual living unrelated donation rate, the black dashed line is the synthetic living unrelated donation rate of New York with all nine weight-assigned states and the gray lines are the leave-one-out estimates.

The average of all nine leave-one-out estimates of the synthetic control (gray lines) are on average 0.1 percent higher than the actual living unrelated donation rate in New York (black solid line) in the pre-legislation period, suggesting that the leave-one-out estimates yield very good fits. Further, the leave-one-out estimates are 0.02 percent lower than the original synthetic New York (black dashed line) in the pre-legislation period. In the post-legislation period, the average of all nine leave-one-out estimates of the synthetic control are on average 4 percent higher than the original synthetic New York. The leave-one-out estimates of the synthetic control are highly robust to the exclusion of any particular state.

3.4.5 Sparse Synthetic Controls

In this section, we create synthetic controls that involve a small number of comparison units. Reducing the number of states in the synthetic control allow us to examine the trade-off between sparsity and goodness of fit in the choice of the number of states that contribute to the synthetic control for New York (Abadie et al., 2014). Accordingly, we construct synthetic controls for New York allowing only combinations of eight, seven, six, five, four, three, two, and a single control state respectively. For $l = 8, 7, 6, 5, 4, 3, 2, 1$, and for all possible combinations of l control states, we choose the one that produces the synthetic control unit that minimizes

[20]The pre-legislation RMSPE is $\left(\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t} - \sum_{j=2}^{s+1} w_j^* Y_{jt}\right)^2\right)^{1/2}$ and the post-legislation RMSPE is $\left(\frac{1}{T_1}\sum_{t=T_0+1}^{T_1}\left(Y_{1t} - \sum_{j=2}^{s+1} w_j^* Y_{jt}\right)^2\right)^{1/2}$ where T_0 and T_1 are the number of pre- and post-legislation periods respectively, $w_j^* Y_{jt}$ is the synthetic outcome using the j^{th} unexposed state with weight w^* and Y_{1t} is the actual outcome of the treated state.

the minimum RMSPE[21].

Table 5 shows the states and weights for the sparse synthetic controls and the compromise in terms of goodness of fit that results by reducing the number of states, l, that contribute to the synthetic control. The states contributing to the sparse versions of the synthetic control for New York are subsets of the set of nine states contributing to the synthetic control in the baseline specification. Nevada has the largest weight in the majority of the cases. Overall, relative to the baseline synthetic control with nine states, the loss in goodness of fit (as shown by the corresponding RMSPE in table 5) is very low for $l = 8, 7, 6, 5, 4, 3$. For the case of $l = 1, 2$, the goodness of fit is significantly poor, showing a substantial gap between the actual and the synthetic living unrelated kidney donation rates.

Figure 4 shows the living unrelated kidney donation rate trajectory for New York and sparse synthetic controls with $l = 8, 7, 6, 5, 4, 3, 2, 1$. The sparse synthetic controls in figure 4 produces results that are highly similar to the original synthetic control shown in figure 1 for cases where $l > 2$. However, for two and one control states, the pre-legislation fits are very poor. This suggests that using combinations of states rather than a single or two comparison units is important in terms of matching pre-legislation kidney donation rates.

4 Discussion

In sum, when we employ methods that do not depend on the assumptions of traditional DiD designs, we find evidence consistent with the possibility that tax incentives encourage unrelated kidney donations. While we are only able to identify this evidence clearly in New York, it is possible that other states' statutes are similarly effective. Again, our methodology does not allow us to say one way or the other. We also cannot say for certain whether the effect we observe is due solely to New York's tax statute, or to some combination of the tax and "medal" statutes.

To the extent New York is relatively unique, that outcome would not be wholly surprising. The mean dollar value we compute for New York's incentive is the highest of any deduction-granting state, at $665 net of federal tax. Our regression results suggest that income levels can affect donations, and New York is one of the highest-earning states. And, in regressions we run only in states enacting some form of incentive legislation, we find that population is strongly and significantly correlated with donations, with a coefficient of about 2.0[22]. In other words, if we were going to find an effect in any state, it would probably be New York.

Six hundred sixty-five dollars is a small sum for undertaking life-threatening surgery, but we think our results are nonetheless plausible. Many would-be donors cite cost as a barrier to donation (Knotts et al., 1996). Financial barriers to donation may be as low as $500, although they may also range into the low five figures (Johnson et al., 1999). Becker and Elias (2007) offer a back-of-the-envelope estimate of nonmonetary costs of about $5,000. The impact of New York's incentive may be not to generate new donors, but to instead to ease these cost barriers for those already inclined to donate. The symbolic and publicity value of the law's enactment may also loom larger than the dollar amount itself (Healy, 2010).

While we hesitate to draw strong conclusions from one state, at a minimum our results suggest the importance of continued research into organ-donor incentives. In 2007, the President's Council on Bioethics

[21]The combination of 8 states out of all 9 weight-assigned states in the baseline specification is $C_8^9 = \frac{9!}{8!(9-8)!} = 9$ meaning that there are 9 set of combinations each consist of 8 control states. Similarly, $C_7^9 = 36$, $C_6^9 = 84$..., so on and so forth. Repeating this exercise for combinations of 5,4,3,2 and 1, we get 8 groups and a total of 510 set of combinations and thus 510 synthetic control estimations. From each group, we pick the one that has the minimum RMSPE which leads to the output in table 5.

[22]Healy (2010) Ch. 3 also reports that population density and income are important determinants of donations.

recommended experiments with a variety of financial incentives for organ donation (Crowe et al., 2007). Momentum for those policies has slowed considerably in the wake of the string of findings that existing incentives have had minimal impact. Yet policies that further expand the pool of donors have enormous potential benefits; for example, even a 5% increase in living donations could add roughly 11,000 quality-adjusted life years for ESRD sufferers (Barnieh et al., 2013).[23]

More generally, our findings provide a cautionary note for those who have argued, following Titmuss (1970), that government-provided financial incentives will tend to crowd out altruistic behavior. The set of circumstances that may have given rise to the appearance of crowding out here could easily appear in other settings. Organizations skilled in motivating their members to work together for altruistic goals may be readily able to translate those skills to political success (Andrews and Edwards, 2004). It is natural to expect that the group's efforts might drop off after legislative agreement to pursue the group's goals (Peltzman, 1973). Our findings suggest that DiD methods may lead to spurious evidence of crowding out, in turn implying that financial incentives may be more efficacious than prior literature suggests.

5 Conclusion

We used state-level panel data for the period 1988-2012 in order to reveal the causal impact of tax incentive legislation on living kidney donation rates in the U.S states. For the sake of completeness and for purposes of cross-validation, we first performed a DiD approach that accounts for several possible dimensions of endogeneity. Then we invoked an instrumental variables approach that accounts for large variations in the value of the incentives offered and the endogeneity of the law. Finally, we employed the synthetic control method that accounts for time-varying unobserved state heterogeneity. The latter method is based on estimating the counterfactual: how the state living related and unrelated kidney donation rates would have evolved in the absence of legislation.

Contrary to prior efforts, we have found evidence consistent with the hypothesis that tax incentives for organ donation increase donations. We argue that failure of the parallel trends assumption gives rise to a spurious negative correlation between enactment and donations. While the DiD and the instrumental variables approaches indicate no statistically significant average causal effects of tax incentive policies on related or unrelated kidney donation rates, the synthetic control method reveals that tax incentives increased living unrelated kidney donation rates in New York by about 52 percent in the post-legislation period (from 2006 to 2012) relative to a comparable synthetic New York in the absence of legislation. We further show that our synthetic control estimate is highly robust to the exclusion of any particular state. Our methods can neither rule out nor confirm that other states' policies were similarly effective. These results may be significant not only for health policy, but also for government efforts to encourage other-regarding behavior generally.

[23] See Cook and Krawiec (2014) for a wider discussion.

References

Abadie, A.; A. Diamond; and J. Hainmueller (2010): Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105:493–505.

Abadie, A.; A. Diamond; and J. Hainmueller (2014): Comparative Politics and The Synthetic Control Method. *American Journal of Political Science (forthcoming)*, p. DOI: 10.1111/ajps.12116.

Abadie, A. and J. Gardeazabal (2003): The Economic Costs of Conflict: A Case Study of the Basque Country. *The American Economic Review*, 93:113–132.

Andrews, K.T. and B. Edwards (2004): Advocacy organizations in the US political process. *Annual Review of Sociology*, 30:479–506.

Barnieh, L.; J.S. Gill; S. Klarenbach; and B.J. Manns (2013): The cost-effectiveness of using payment to increase living donor kidneys for transplantation. *Clinical Journal of the American Society of Nephrology*, pp. doi: 10.2215/CJN–03350313.

Becker, G.S. and J.J. Elias (2007): Introducing incentives in the market for live and cadaveric organ donations. *The Journal of Economic Perspectives*, 21(3):3–24.

Bertrand, M.; E. Duflo; and S. Mullainathan (2004): How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Bilgel, F. (2013): The effectiveness of transplant legislation, procedures and management: Cross-country evidence. *Health Policy*, 110:229–242.

Bilgel, F. (2014): The Effect of Paid Leave of Absence Legislation on Living Kidney Donation in the United States. In: *Mimeo, Okan University*.

Boulware, L.E.; M.U. Troll; L.C. Plantinga; and N.R. Powe (2008): The association of state and national legislation with living kidney donation rates in the United States: A National Study. *American Journal of Transplantation*, 8(7):1451–1470.

Calandrillo, S. (2004): Cash for Kidneys? Utilizing Incentives to End America's Organ Shortage. *George Mason Law Review*, 13:69–133.

Cameron, A.C.; J.B Gelbach; and D.L Miller (2011): Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29(2):238–249.

Cook, P.J. and K.D. Krawiec (2014): A Primer on Kidney Transplantation: Anatomy of the Shortage. *Law & Contemporary Problems (forthcoming)*, 77.

Crowe, S.; E. Cohen; and A. Rubenstein (2007): *Increasing the supply of human organs: Three Policy Proposals*. President's Council on Bioethics.

Finlay, K.; L.M. Magnusson; and M.E. Schaffer (2013): *weakiv: Weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models*. Tech. rep.

Gneezy, U.; S. Meier; and P. Rey-Biel (2011): When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, pp. 191–209.

Healy, K. (2010): *Last best gifts: Altruism and the market for human blood and organs.* University of Chicago Press.

Johnson, E.M.; J.K. Anderson; C. Jacobs; G. Suh; A. Humar; B.D. Suhr; S.R. Kerr; and A.J Matas (1999): Long-Term Follow-Up of Living Kidney Donors: Quality of Life After Donation. *Transplantation*, 67(5):717–721.

Knotts, R.S.; W.F. Finn; and T. Armstrong (1996): Psychosocial factors impacting patients, donors, and nondonors involved in renal transplant evaluation. *Kidney Found Perspect*, 15:11–23.

Lacetera, N.; M. Macis; and S. S. Stith (2014): Removing Financial Barriers to Organ and Bone Marrow Donation: The Effect of Leave and Tax Legislation in the U.S. *Journal of Health Economics*, pp. 43–56.

Matas, A.J. and M. Schnitzler (2004): Payment for Living Donor (Vendor) Kidneys: A Cost-Effectiveness Analysis. *American Journal of Transplantation*, 4(2):216–221.

Moreira, J.M. (2003): A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71:1027–1048.

Olson, M. (1971): *The logic of collective action: Public goods and the theory of groups.* Harvard University Press.

Peltzman, S. (1973): The effect of government subsidies-in-kind on private expenditures: The case of higher education. *The Journal of Political Economy*, pp. 1–27.

Satel, S.L. and B.E. Hippen (2007): When Altruism Is Not Enough: The Worsening Organ Shortage and What It Means for the Elderly. *The Elder Law Journal*, 15:153.

Schaffer, M.E. (2010): *xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models.* Tech. rep., Heriot-Watt University.

Stock, J. H. and M. Yogo (2005): *Testing for Weak Instruments in Linear IV Regression. In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.* Cambridge University Press.

Titmuss, R. (1970): *The gift relationship: From human blood to social policy.* Allen and Unwin London.

Venkataramani, A.S.; E.G. Martin; A. Vijayan; and J.R. Wellen (2012): The impact of tax policies on living organ donations in the United States. *American Journal of Transplantation*, 12(8):2133–2140.

Wellington, A.J. and E.A. Sayre (2011): An evaluation of financial incentives policies for organ donations in the United States. *Contemporary Economic Policy*, 29:1–13.

A Tables

Table 1: Difference-in-differences Estimates, 1988-2012

Outcome variable	Living related kidney donation rate pmap			Living unrelated kidney donation rate pmap		
	LIML	CGM-LIML	Residual aggregation	LIML	CGM-LIML	Residual aggregation
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-	-	-20.226 (15.038)	-	-	-1.626 (5.874)
Tax incentive	-1.913 (1.954)	-1.913 (10.887)	-0.326 (0.373)	0.098 (0.822)	0.098 (0.826)	0.028 (0.113)
Kidney waiting list additions, pmp	0.183***(0.035)	0.183 (0.241)	0.185***(0.019)	0.031***(0.004)	0.031 (0.021)	0.034***(0.003)
Cadaveric donation rate, pmap	-0.290**(0.119)	-0.290 (2.830)	-0.117**(0.050)	-0.104 (0.091)	-0.104 (0.108)	-0.063*(0.033)
Prevalence of ESRD pmp	0.035**(0.015)	*partialled-out*	0.029***(0.007)	-0.003 (0.003)	*partialled-out*	-0.002 (0.003)
Number of transplant centers pmap	9.488**(4.473)	9.488 (43.070)	8.494**(3.172)	-0.410 (2.444)	0.410 (2.748)	1.564 (0.945)
GDP per capita	-0.0002 (0.0004)	-0.0002 (0.002)	-0.0001 (0.0005)	-0.0001 (0.0001)	-0.0001 (0.0002)	-0.00005 (0.0001)
R-squared	0.5349	0.3710	0.7498/0.0358	0.6822	0.1366	0.7965/0.0009
Number of observations	624	624	663/28	624	624	663/28
Endogeneity test (p-value)	2.115 (0.1459)	-	-	0.081 (0.7754)	-	-
Underidentification test (p-value)	9.267 (0.0548)	-	-	4.634 (0.0986)	-	-
Weak identification test	13.563	13.563	-	23.042	23.042	-
Conditional likelihood ratio (p-value)	1.25 (0.3274)	6.38 (0.1691)	-	0.00 (0.9576)	-	-
Sargan-Hansen statistic (p-value)	4.509 (0.2115)	4.258 (0.2350)	-	1.219 (0.2696)	0.689 (0.4064)	-
Difference-in-Sargan statistic (p-value)	2.95 (0.0860)	6.43 (0.0112)	-	1.77 (0.1831)	0.16 (0.6892)	-
Bootstrap replications	-	2000	-	-	2000	-

Note: All specifications include state and year fixed-effects and state linear trends. In columns (1) and (2), the cadaveric donation rate is instrumented by cerebrovascular deaths, motor vehicle fatalities and their squares. In columns (4) and (5), the cadaveric donation rate is instrumented by cerebrovascular deaths and motor vehicle fatalities. In columns (2) and (5), the prevalence of ESRD, state linear trends and year fixed-effects are partialled out to render the estimated covariance matrix moment conditions full rank. The endogeneity test reports the chi-square and the p-value for the null hypothesis that the cadaveric donation rate is exogenous. The underidentification test reports the Kleibergen-Paap rk LM statistic and the p-value for the null hypothesis that the equation is underidentified (the excluded instruments are irrelevant). The weak identification test reports the Cragg-Donald Wald F-statistic for the null hypothesis that the equation is weakly identified. Stock and Yogo (2005) weak identification test critical values for 10% maximal LIML size are 8.68 in columns (4) and (5) and 5.44 in columns (1) and (2). Moreira (2003)'s conditional likelihood ratio reports the weak identification-robust inference likelihood ratio and the p-value for the null hypothesis that the coefficient on the cadaveric donation rate is zero. Sargan-Hansen statistic reports the chi-square and the p-value for the joint null hypothesis that the instruments are uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimated equation (i.e. the instruments are valid). The difference-in-Sargan test statistic reports the chi-square and the p-value for the null hypothesis that the equation is well specified and has no neglected nonlinearities. Two-way cluster-robust bootstrapped standard errors in parentheses in columns (2) and (5). Arbitrary heteroscedasticity and autocorrelation robust standard errors in columns (1) and (4). $***P < .01$ $**P < .05$ $*P < .10$

Table 3: Synthetic Control States

Treated States	Donor States		
Louisiana	Alabama	North Carolina	Nevada
Rhode Island	Arizona	Nebraska	South Dakota
New York	Florida	New Hampshire	Tennessee
	Kentucky	New Jersey	Vermont
	Michigan		

Table 4: Synthetic Control Weights for New York, Living Unrelated Kidney Donation

State	Weight
Alabama	0.179
Arizona	0.177
Florida	0.107
Michigan	0.074
Nebraska	0.020
New Jersey	0.058
Nevada	0.269
Tennessee	0.089
Vermont	0.027

Table 2: Limited Information Maximum Likelihood (LIML) Estimates, 1997-2011

Outcome variable	Natural log of living related kidney donation rate pmap				Natural log of living unrelated kidney donation rate pmap			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean value of state donation incentive	-0.098 (0.060)	-0.089 (0.064)	-0.110 (0.069)	-0.044*(0.025)	0.554 (5.562)	0.480 (0.978)	0.345 (0.827)	-0.037 (0.049)
Kidney waiting list additions, pmp	0.540***(0.166)	0.472***(0.162)	0.467***(0.158)	0.434***(0.094)	0.725 (2.123)	0.875 (0.638)	0.830 (0.536)	0.407***(0.139)
Cadaveric donation rate, pmap	0.762 (0.505)	0.883 (0.543)	0.890*(0.514)	0.104 (0.098)	-0.237 (3.244)	-0.324 (0.868)	-0.193 (0.696)	-0.191 (0.188)
Prevalence of ESRD pmp	0.556 (1.061)	0.488 (1. 140)	-0.721 (1.551)	-0.923 (1.003)	4.412 (14.687)	4.496 (7.126)	4.213 (7.145)	-0.010 (1.686)
Number of transplant centers pmap	0.296 (0.198)	0.338 (0.236)	0.347 (0.239)	0.511***(0.174)	-0.467 (14.236)	-0.401 (2.935)	-0.004 (2.331)	0.075 (0.345)
GDP per capita	0.690 (1.006)	-	-	-0.644 (0.706)	2.372 (15.519)	-	-	-0.790 (0.867)
Health spending	-	0.051 (0.123)	0.005 (0.149)	-	-	0.761 (1.036)	0.652 (0.910)	-
Unemployment rate	-	-	-0.613*(0.335)	-0.408**(0.175)	-	-	-0.387 (1.885)	-0.579**(0.291)
Share of itemizers	-	-	-2.321 (1.759)	-0.786 (0.629)	-	-	2.585 (6.576)	-0.718 (0.858)
Median Income (2005)	-	-	-0.449 (0.563)	-0.233 (0.433)	-	-	-2.427 (2.011)	-0.740 (0.567)
Total population	-	-	-	-1.521 (1.041)	-	-	-	-0.686 (1.569)
Share of black population	-	-	-	-0.008 (0059)	-	-	-	-0.233**(0.111)
State medicaid spending	-	-	-	-0.053 (0.180)	-	-	-	-0.264 (0.245)
Paid leave for public employees	-	-	-	0.085 (0.056)	-	-	-	0.009 (0.099)
Paid leave for private employees	-	-	-	-0.129*(0.077)	-	-	-	-0.100 (0.133)
Number of observations	532	494	494	377	532	494	494	367
Underidentification test (p-value)	10.329 (0.0352)	9.160 (0.0572)	10.108 (0.0386)	10.932 (0.0121)	10.329 (0.035)	9.160 (0.0572)	10.108 (0.0386)	12.288 (0.0065)
Weak identification test	5.610	4.878	4.705	3.964	5.610	4.878	4.705	4.466
Conditional likelihood ratio (p-value)	1.57 (0.2848)	1.14 (0.3635)	1.38 (0.3011)	4.17 (0.0767)	4.85 (0.0627)	3.76 (0.0892)	2.57 (0.1681)	0.12 (0.7510)
Hansen J statistic (p-value)	3.556 (0.3136)	3.359 (0.3394)	2.776 (0.4274)	1.961 (0.3751)	2.448 (0.4847)	2.834 (0.4179)	3.983 (0.2633)	3.442 (0.1789)
Difference-in-Sargan statistic (p-value)	0.12 (0.7323)	0.30 (0.5853)	0.62 (0.4318)	4.19 (0.0408)	1.26 (0.2619)	1.83 (0.1762)	1.58 (0.2083)	0.08 (0.7725)

Note: All explanatory variables are expressed as natural logarithm. All specifications include state and year fixed-effects. In columns 1-3 and 5-7, the natural log of mean value of state incentive for donation is instrumented by the IIT marginal rate, the federal medicaid matching grant to state and their squares. In columns 4 and 8, the mean value of state incentive for donation is instrumented by the IIT marginal rate, its square and the interaction of ESRD trend and the average federal marginal rate. The underidentification test reports the Kleibergen-Paap rk LM statistic and the p-value for the null hypothesis that the equation is underidentified (i.e. the excluded instruments are irrelevant). The weak identification test reports the Cragg-Donald WaldF-statistic for the null hypothesis that the equation is weakly identified. Stock and Yogo (2005) weak identification test critical values for 10% and 15% maximal LIML size are 5.443.87 respectively. Moreira (2003)'s conditional likelihood ratio reports the weak identification-robust inference likelihood ratio and the p-value for the null hypothesis that the coefficienton the mean value of state incentive for donation is zero. Hansen J statistic reports the chi-square and the p-value for the joint null hypthesis that the instruments are uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimated equation (i.e. the instruments are valid). The difference-in-Sargan test statistic reports the chi-square and the p-value for the null hypothesisthat the equation is well specified and has no neglected nonlinearities. Standard errors in parenthesis are robust to arbitrary heteroscedasticity and autocorrelation.

$***P < .01$ $**P < .05$ $*P < .10$

Table 5: Synthetic Weights from Combination of Control States, New York

Synthetic		State and Weights								
Combination	RMSPE	MI	NV	AL	AZ	TN	FL	VT	NJ	NE
Nine control states	0.1658	0.074	0.269	0.179	0.177	0.089	0.107	0.027	0.058	0.020
Eight control states	0.1681	0.088	0.287	0.194	0.172	0.091	0.093	0.029	0.045	
Seven control states	0.1693	0.058	0.316	0.201	0.194	0.095	0.108	0.027		
Six control states	0.1760	0.107	0.322	0.182	0.200	0.077	0.111			
Five control states	0.1854	0.106	0.418	0.209	0.183	0.083				
Four control states	0.2072	0.129	0.433	0.257	0.180					
Three control states	0.2619	0.202	0.428	0.370						
Two control states	0.5122	0.706	0.294							
One control state	1.0500	1.000								

Note: States and weights are constructed from the best fitting combination of nine, eight, seven, six, five, four, three, two and one states.

B Figures

Figure 1: Living unrelated kidney donation rate trajectories: New York vs. synthetic New York

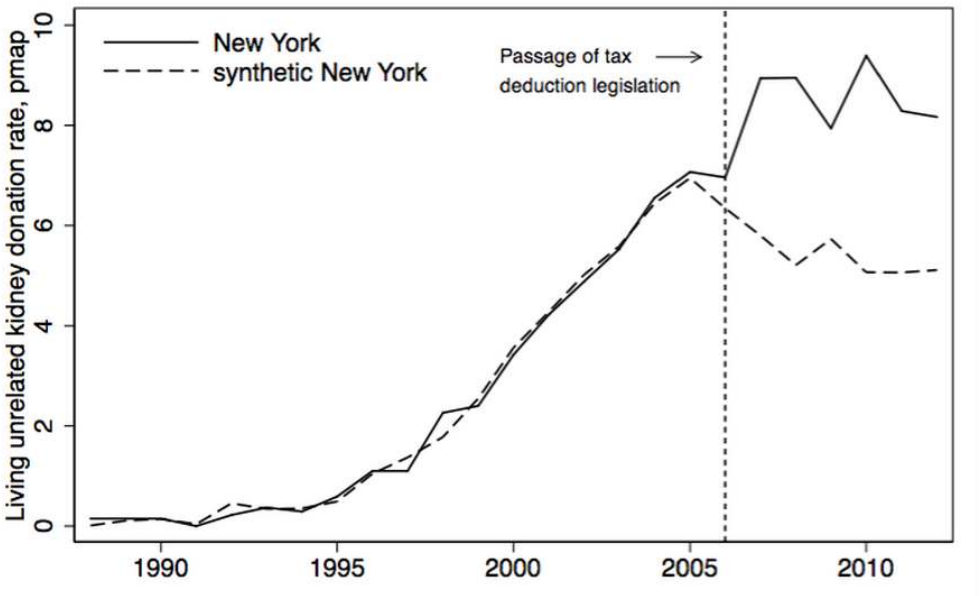


Figure 2: In-space placebo distributions, living unrelated kidney donation

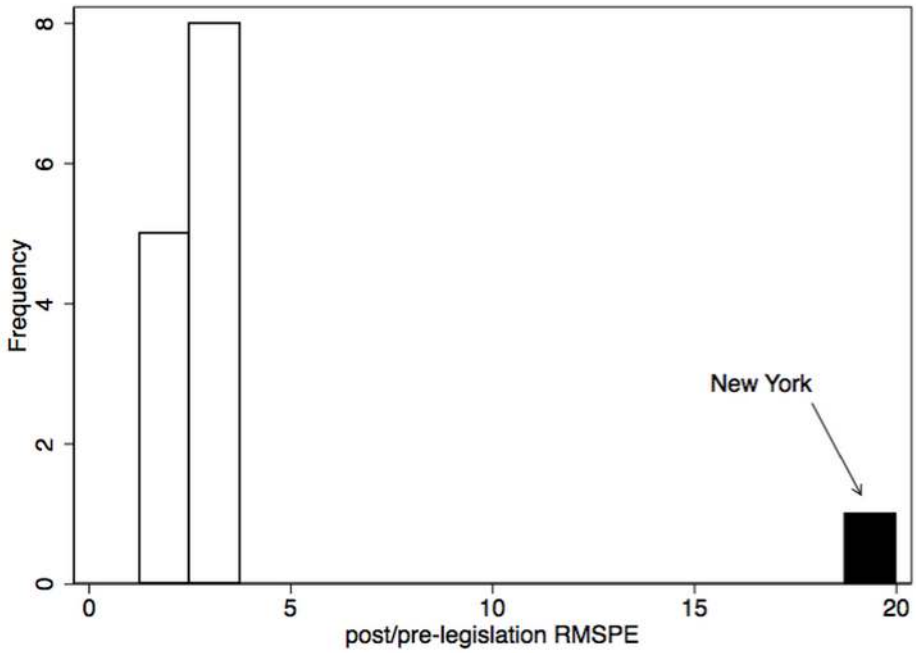


Figure 3: Leave-one-out distribution of the synthetic control, living unrelated kidney donation

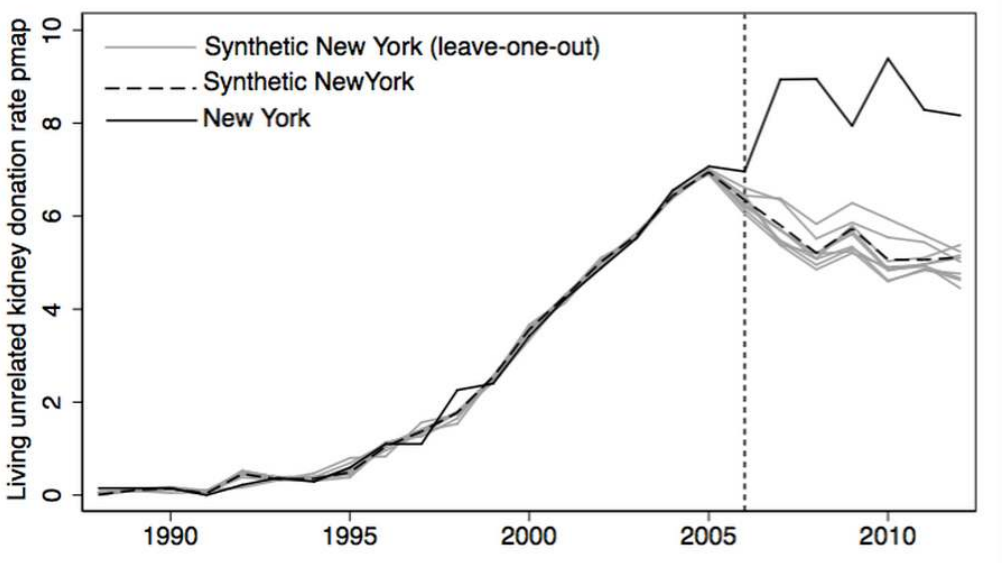
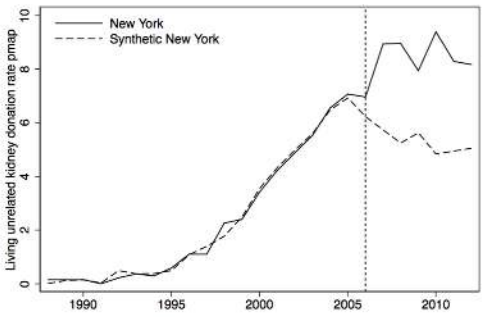
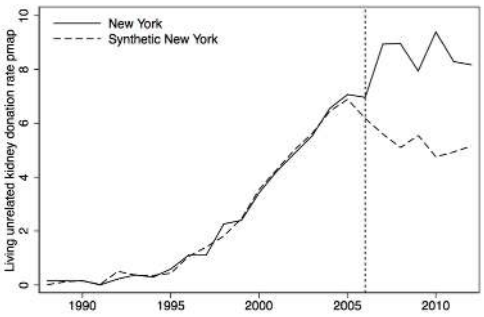


Figure 4: Living unrelated kidney donation rate gaps between New York and sparse synthetic controls

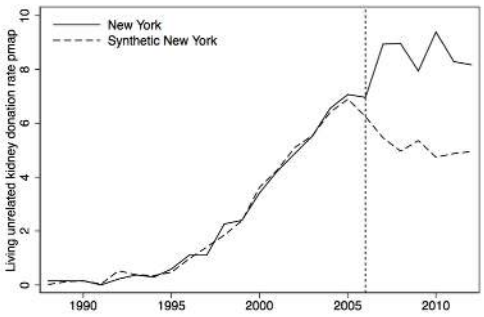
(a) Number of control states: 8



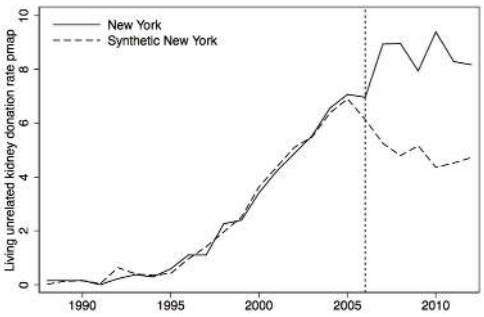
(b) Number of control states: 7



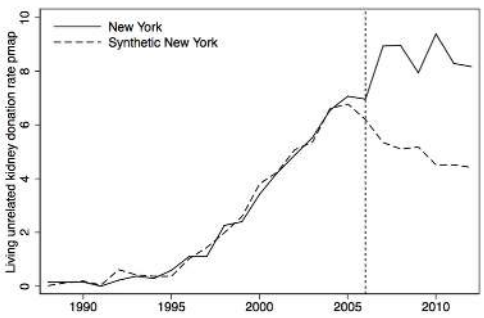
(c) Number of control states: 6



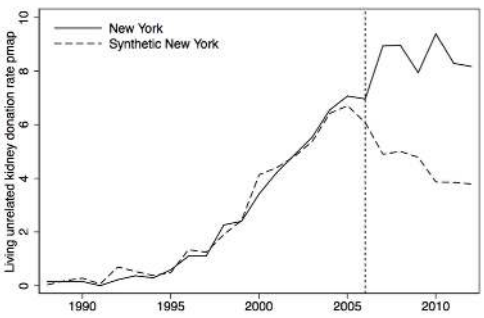
(d) Number of control states: 5



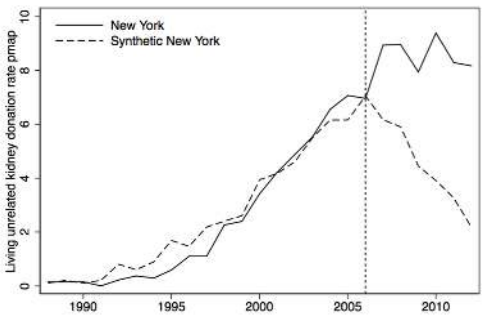
(e) Number of control states: 4



(f) Number of control states: 3



(g) Number of control states: 2



(h) Number of control states: 1