DOCUMENTS DE TRAVAIL i3 – i3 WORKING PAPERS

interdisciplinary institute on innovation

Methods for strengthening a weak instrument in the case of a persistent treatment.

Michel Berthélémy
MINES ParisTech, PSL – Research University, CERNA – Centre for industrial economics, Paris
michel.berthélémy(a)mines-paristech.fr

Petyo Bonev
MINES ParisTech, PSL – Research University, CERNA – Centre for industrial economics, Paris
petyo.bonev(a)mines-paristech.fr

Damien Dussaux
MINES ParisTech, PSL – Research University, CERNA – Centre for industrial economics, Paris
damien.dussaux(a)mines-paristech.fr

Magnus Söderberg
MINES ParisTech, PSL – Research University, CERNA – Centre for industrial economics, Paris
magnus.soderberg(a)mines-paristech.fr

Working Paper 15-CER-04
April 2, 2015

Pour citer ce papier / How to cite this paper:

Bethélémy, M., Bonev, P., Dussaux, D., Söderberg, M. (2015) Methods for strengthening a weak instrument in the case of a persistent treatment. i3 Working Papers Series, 15-CER-04.

Methods for strengthening a weak instrument in the case of a persistent treatment

Michel Berthélemy,[a] Petyo Bonev,[a,b] Damien Dussaux [a] and Magnus Söderberg [a]

[a] MINES ParisTech, PSL Research University, CERNA – Centre d'Economie Industrielle, 60 Boulevard St Michel, 75006 Paris, FRANCE.

[b] Corresponding author: E-mail: petyo.bonev@mines-paristech.fr. Tel: +33 (0)1 40 51 91 85. Fax: +33 (0)1 40 51 91 45.

31 March 2015

Abstract

When evaluating policy treatments that are persistent and endogenous, available instrumental variables often exhibit more variation over time than the treatment variable. This leads to a weak instrumental variable problem, resulting in high bias or uninformative confidence intervals. We propose two new estimation approaches that strengthen the instrument. We derive their theoretical properties and show in Monte Carlo simulations that they outperform standard IV-estimators. We use our procedures to estimate the effect of public utility divestiture in the U.S. nuclear energy sector. Our results show that divestiture significantly increases production efficiency.

Keywords: weak instrument, treatment evaluation, nuclear power

JEL codes: C22, C26, L25, L94

1. Introduction

We consider a situation where analysts have access to panel observations of a binary policy variable (treatment) and a dependent variable (policy outcome). It is assumed that the treatment is endogenous and that an instrumental variable is available. Moreover, in a context of a policy evaluation, the treatment is sometimes persistent, i.e. once the policy is implemented it remains in place for several or all remaining periods.[1] Leading examples of such policies are legal and regulatory changes in infrastructure sectors where assets have long lifetimes. Moreover, typical instruments are based on economic shocks and exhibit much higher variation over time than the treatment variable.[2] As a result, the instrument becomes weak and, furthermore, it weakens over time even when it is strong on the cross sectional level. We refer to this phenomenon as the persistent treatment problem. In this paper, we propose two methods that strengthen the instrument by excluding unnecessary variation.

The first method, which we denote as the Forward Variation Reduction (FVR) approach, takes the value of the instrument in the first treated period and copies it to all future periods. This transformation is carried out for each unit. Next, a standard Two Stage Least Squares (TSLS) is performed using the transformed instrument. The intuition behind the FVR approach is that variation in the instrumental variable is uninformative in periods after the implementation of the treatment due to the persistence of the treatment.

The second method, the Forward and Backward Variation Reduction (FBVR) approach, is an extension of the FVR approach. As its name indicates, the instrument variation is also

[1] Policy persistence can occur because: 1) it takes time to evaluate a policy change since, for example, policy shifts obstruct information about true market conditions (Warren and Wilkening, 2012); 2) it might not be possible to implement another change quickly since market agents may lobby for the protection of sunk investments (Coate and Morris, 1999); 3) uncertainty about future gains and losses alters voters' preferences in favor of the status quo (Fernandez and Rodrik, 1991).

[2] Examples and references are provided in Section 2.

restricted backwards: all instrument values during periods prior to the period when the treatment is first implemented are set equal to the value of the instrument in the last untreated period. In this way, we only retain the variation that triggers a change in the treatment variable. For both approaches, we prove identification and derive asymptotic properties.

In Monte Carlo simulations, we study the small sample properties of the FVR and FBVR approaches and compare them to OLS and standard TSLS. We construct a data generation process that allows us to compare the results obtained at different instrument strength, holding the endogeneity level constant, and vice versa. The simulations build on a discrete choice framework that has been used earlier by Honoré (2006) and Carro (2007). The FVR and FBVR approaches perform substantially better than both OLS and TSLS: FVR/FBVR 95% confidence intervals are up to 70 times smaller than the TSLS empirical standard errors and the empirical bias is up to 10 times smaller than the empirical OLS bias. These results are robust to choices of instrument distribution, endogeneity level and instrument strength.

In addition to their superior small sample performance, the FVR and FBVR have various conceptual and practical advantages. First, they can be applied in a nonlinear panel data setting, as the crucial idea only relies on the separability of the unobserved idiosyncratic error term. Second, both approaches are easy to implement, intuitive and therefore readily accessible to practitioners. Third, in comparison to first differencing, where only the last untreated and first treated observations are used, no observations are ignored. We demonstrate the importance of the last point in a simulation study.

Lastly, we use the FVR/FBVR approaches to evaluate the effect of public utility divestiture on nuclear reactor production availability in the U.S. To the best of our knowledge, this is the first empirical study that takes both the persistence and endogeneity of the treatment into consideration. We find that divestiture of U.S. nuclear reactors causes a significant increase in their production availability by at least 7%.

Section 2 formally describes the problem of persistent treatment and relates it to the relevant literature. Section 3 defines the FVR and FBVR approaches and explains how they strengthen a weak instrument in the case of a persistent treatment. Section 4 evaluates the small sample properties of the FVR/FBVR and alternative estimators that are common in the literature using Monte Carlo simulations. Section 5 applies the methods to evaluate the effect of divestiture on U.S. nuclear reactors' operating performance. Section 6 concludes.

2. The problem of persistent treatment

Suppose there are panel observations on a binary random variable D_{it} (treatment) and on an outcome variable Y_{it}. Index i indicates the cross sectional unit, where $i = 1, \ldots, n$, and index $t = 1, \ldots, T$ indicates the time period. As motivating examples, D_{it} might be an indicator variable for market deregulation or for obtaining a college degree. Y_{it} might be a measure of firm production efficiency or individual wages. In many cases, the treatment variable D_{it} is potentially endogenous due to unobserved selection of units into (or out of) treatment. We consider a period-specific instrument Z_{it}, for D_{it}. Using the exogenous variation of the Z_{it}, it is often possible to identify the causal effect of D_{it} on Y_{it}, see e.g. Angrist and Krueger (2001). The persistent treatment problem arises when the variation of the instrument over time is much higher than the variation of the treatment variable. In particular, the following features lead to a persistent treatment problem:

- The treatment is endogenous,

- The treatment is persistent, i.e. once a unit is treated, the treatment variable does not change its value for many, or all, subsequent periods,
- There is an instrument whose values vary from period to period.

These features are common when policies are evaluated. Instruments often vary more than policy-state variables over time since instruments are frequently based on economic shocks. Examples of such instruments are source-weighted exchange rates (Revenga, 1990, 1992; Bertrand, 2004) and exposure to oil shocks (Raphael and Winter-Ebner, 2001). As a result, the greater the number of periods, the weaker the instrument.

To formalize the problem, we assume the standard fixed effects linear model,

$$Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}, \tag{1}$$

where X_{it} is a $1 \times K$ dimensional random vector of observed individual characteristics, C_i is unobserved and time-constant, U_{it} is the unobserved error term and α is the coefficient of primary interest.

We allow for two types of endogeneity. First, $corr(X_{it}, C_i)$ is not necessary zero. Typical examples of C_i are firm culture and management quality. Firm culture might be correlated with observable expenditure for maintenance, which is captured by X_{it}. Second, $corr(D_{it}, U_{it})$ is not necessarily zero. U_{it} might capture an anticipation of a tax reduction so that it correlates with firm performance. Similarly, U_{it} might capture unobserved work effort exerted by individual *i* that correlates with the intention/effort to obtain specialized education. Due to the endogeneity of the treatment variable, the standard fixed effects (FE) estimator is potentially biased. Assume further, that there is an observable *M*-dimensional random vector $\mathcal{L}_{it} = \left(\mathcal{L}_{it,1},\ \mathcal{L}_{it,2}, \ldots,\ \mathcal{L}_{it,M}\right)$ that is exogenous and can be used as an instrument for the endogenous treatment. We write $W_{it} \coloneqq (D_{it}, X_{it})$ and $Z_{it} \coloneqq (X_{it}, \mathcal{L}_{it})$. Furthermore, we define $\bar{Y}_i \coloneqq \frac{1}{T}\sum_{t=1} Y_{it}$ and $\tilde{Y}_{it} \coloneqq Y_{it} - \bar{Y}_i$ (and with analogous notation for all other random variables). The demeaned model is

$$\tilde{Y}_{it} = \alpha \tilde{D}_{it} + \tilde{X}_{it}\beta + \tilde{U}_{it}, \tag{2}$$

or, equivalently,

$$\tilde{Y}_{it} = \tilde{W}_{it}\varphi + \tilde{U}_{it}, \tag{3}$$

where $\varphi = (\alpha, \beta')'$. Finally, using matrix notation, model (3) can be written as

$$\tilde{Y}_i = \tilde{W}_i\varphi + \tilde{U}_i, \tag{4}$$

where $\tilde{Y}_i = \left(\tilde{Y}_{i1},\ \tilde{Y}_{i2}, \dots,\ \tilde{Y}_{iT}\right)$ and analogously for $\tilde{W}_i$ and $\tilde{U}_i$. The standard approach is to use a pooled TSLS method to estimate φ. The standard assumptions are:

1: $\mathbb{E}[U_{it}|Z_{i1}, Z_{i2}, \dots, Z_{iT}, C_i] = 0$ for $t = 1, \dots, T$.

2: $rank\left(\mathbb{E}[\tilde{Z}'_{it}\tilde{Z}_i]\right) = \mathrm{rank}\left(\sum_{t=1}^{T} \mathbb{E}[\tilde{Z}'_{it}\tilde{Z}_i]\right) = L$, where L is the dimension of $Z_{it}, L = K + M$

3: $rank\left(\mathbb{E}[\tilde{Z}'_{it}\tilde{X}_i]\right) = \mathrm{rank}\left(\sum_{t=1}^{T} \mathbb{E}[\tilde{Z}'_{it}\tilde{X}_i]\right) = K$.

The important assumptions in this study are assumptions 1 and 3. Assumption 1 states the strict exogeneity assumption. Assumption 3 is a rank condition that states that the instrument and the endogenous regressor are sufficiently related.

In this context, persistence of the treatment may cause the following problems. First, it might lead to a violation of the second rank condition (assumption 3). As a result, the causal effect would not be identified. Intuitively, if Z_{it} and X_{it} are not related, then variation in Z_{it} cannot be used to reveal the causal parameter, even with an infinitely large sample of observations (Y_i, X_i, Z_i). Second, even if the parameter is identified, persistence of the treatment could lead to estimation problems due to the weak instrument. These problems are well known in the literature on weak instruments in a TSLS context (Stock et. al., 2002). In the just-identified case, the asymptotic variance is potentially very high, implying that the confidence intervals are uninformative. Moreover, the confidence intervals may not have the correct nominal coverage (Staiger and Stock, 1997). The general literature on weak instruments has focused mainly on achieving the correct confidence intervals, starting with Anderson and Rubin's pioneering paper in 1949. One study that considers strengthening the instrument is Ratkovic and Shirato (2014). This study considers that the instrument is weak because some agents are not influenced by the instrument (non-compliers). The authors tackle the problem by down-weighting those observations. In our case, however, the weakness of the instrument evolves over time due to the persistence of the treatment and this method is not applicable.

In the over-identification case with weak instruments, TSLS might be severely biased and inconsistent, and normal approximations may lead to a dramatic understatement of the width of confidence intervals, (Staiger and Stock, 1997; Hahn and Hausman, 2003). These problems can occur in situations that are highly relevant to empirical work, see e.g. Bound et al. (1995). In such cases, the limited information maximum likelihood (LIML) is substantially less biased than the TSLS and can be adapted to produce confidence intervals with the correct nominal coverage (Bekker, 1994; Flores-Lagunes, 2007). In applications, however, many instruments might not be available, or, as in the case of instruments interacting with other exogenous variables, the information from additional instruments might be so limited that the confidence intervals remain very wide and highly uninformative about the sign of the effect. The focus of this paper is on the just-identified (or slightly over-identified) case: we consider a binary, endogenous treatment variable and a single (or a few) instrument(s).

A distinctive characteristic of the persistent treatment problem is that the instrument becomes weak over time. Conditional on treatment in some period t_0 the variables D_{it} and the instrument $\mathcal{L}_{it}$ are independent for all $t > t_0$ Thus, even if the instrument is strong on a cross-sectional level, it becomes weak on a panel level due to the persistence of the treatment. The higher the number of time periods, the more severe the problem. Note that the persistent treatment problem does not rely on linearity of model (1). Any model possessing the three features listed above will induce a weak instrument problem.

Based on these considerations, potential solutions are: (*i*) transform the instrument to obtain less variation, (*ii*) change the structure of the panel, or (*iii*) redefine the treatment and/or the model. One possibility for (*i*) is to set the values of the instrument after the treatment so as to be equal to its value in the first treated period. One possibility for (*ii*) is to restrict the sample of observations by considering a smaller number of periods. Excluding observations before and after the treatment for each cross-sectional unit potentially ensures the rank condition in the new data set. There are several possibilities for (*iii*). First, it might be possible to set the research question in a dynamic framework by defining a dynamic treatment effect. Often the effect of a policy reform does not occur directly after its implementation but over a longer period of time. In such cases α would be a function of time, $\alpha(t)$. Then, the object of interest could be the treatment effect for a given period of time, $\alpha \coloneqq \alpha(t_0)$ for a fixed t_0. This kind of dynamic matching estimators has been proposed by Sianesi (2004) and Fredriksson and Johansson (2008). Typically, these papers rely on the strong assumption that dynamic selection is driven solely by unobservables. This assumption is easily violated in many applications if the data is not rich enough. In addition, inference with these models is often not possible because the asymptotic theory is very demanding and still not fully developed, as for example in Fredriksson and Johansson (2008). Second, there is the literature on dynamic discrete choice models, e.g. Taber (2000) and Heckman and Navarro (2007). In each of finitely many consecutive periods, an agent can choose to take or refuse the treatment. Identification relies on the (semi-)parametric structure, period-specific exclusion restriction and the so called identification-at-the-limit assumption, typical in discrete choice models. In particular, the selection in each period is modeled and the reduced form parameters are identified by conditioning them on the very high values of the instrument in the preceding choices, where the treatment is almost always chosen. However, the large support condition is often not satisfied in applications, for example when the instrument is discrete or hard to justify. In addition, due to their complexity, these approaches are often not accessible to practitioners.

In this paper, we approach the problem by reducing the variation of the instrument. It is accessible to practitioners and its asymptotic properties are technically easy to analyze.

3. Two ways of reducing instrument variation

In this section we first describe the two approaches for strengthing the instrument. Next, we discuss identification and finally, we analyze the finite sample and asymptotic properties.

3.1. The FVR approach

The FVR approach consists of two steps. First, the instrument is transformed in the following way. For each cross-sectional unit, the values of the instrument for all treatment periods are set equal to the value of the instrument in the first treatment period. In the second step, TSLS is performed with the transformed instrument. The simple example in Table 1 illustrates this procedure. The second column contains the dependent variable, for example the percentage of operating hours of a nuclear reactor during. The third column contains the values of the treatment variable, e.g. a dummy variable representing market deregulation. The fourth column contains the instrument, e.g. the number of lobby group members. The last column contains the values of the transformed instrument. The treatment is received in period 3, and the value of the instrument in this period is copied to all subsequent periods (here only to period 4) to create the transformed instrument.

Table 1. FVR, data example

Period	***Y***	***D***	***Z***	$\boldsymbol{Z_{FVR}}$
1	65%	0	14,295	14,295
2	64%	0	13,700	13,700
3	70%	1	15,487	15,487
4	72%	1	12,001	15,487

The intuition behind this procedure is the following. Once the treatment is implemented, the variation of the instrument becomes uninformative. The FVR approach removes this variation and the instrument becomes stronger.

For a formal definition, we introduce the following additional notation. Let T_i be the period at which the agent i is treated for the first time, i=1,..., n and let $T_i > 1$ for all i. Let $P_i = P(T_i)$ denote the random $T \times T$ transformation matrix that is defined in the following way. Its first $T_i - 1$ columns are equal to the first $T_i - 1$ columns of the $T \times T$-identity matrix I_T. Its T_i column has 1 as elements p_{kl} for which $k \geq l$ and 0 elsewhere. The columns $T_i + 1, \dots, T$ consist entirely of zeros. For units with $T_i > T$, we set $P_i = I_T$. That is, the values of the instrument in these cases remain unchanged. For such units, the values of Z_{it}, W_{it}, Y_{it} for $t > T$ and the value of T_i are not observed. In this case, T_i might be either finite (censored T_i) or infinite (non-treated). For simplicity, we focus on the first case, that is, we assume that T_i can be at most equal to $\bar{T} < \infty$. This restriction can easily be relaxed and has no influence on the main intuition and results. As an example for P_i, suppose that $T = 4$ and that for some i, $T_i = 3$. Then $P_i = P_i(3)$ is equal to

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Multiplication by $P(T_i)$ transforms the vector $(a_1, \ldots, a_T)'$ into the vector $(a_1, \ldots\ a_{T_i-1}, a_T, a_T, \ldots, a_{T_i})'$. Furthermore, define the (deterministic) time-demeaned matrix Q_T as

$$Q_T = I_T - j_T(j_T{}'j_T)^{-1}j_T{}',$$

where j_T is a $T \times 1$-vector with all elements equal to 1. Multiplication by Q_T transforms the vector $(a_1, \ldots, a_T)$' into the vector $(a_1 - T^{-1}\sum_{t=1}^{T} a_t\,, \ldots, a_T - T^{-1}\sum_{t=1}^{T} a_t)'$.

The forward variation reduction estimation approach can now be described in the following way.

Step 1: Replace $\mathcal{L}_i$ with $\ddot{\mathcal{L}}_i \coloneqq P_i\mathcal{L}_i$. Write $\ddot{Z}_i \coloneqq (X_i, \ddot{\mathcal{L}}_i)$ and let the rows of the matrix $\ddot{Z}_i$ be denoted by $\ddot{Z}_{it} = (X_{it}, \ddot{\mathcal{L}}_{it})$, $t = 1, \ldots, T$.

Step 2: Estimate the equation

$$Q_T Y_i = Q_T W_i \gamma + Q_T U_i \qquad (5)$$

via (pooled) TSLS using $Q_T\ddot{Z}_i$ as an instrument for $Q_T W_i$. The estimator is defined as

$$\hat{\gamma}_{FVR} \coloneqq \left[\left(\sum_{i=1}^{n}\widetilde{W}_i{}'\tilde{\ddot{Z}}_i\right)\left(\sum_{i=1}^{n}\tilde{\ddot{Z}}_i{}'\,\tilde{\ddot{Z}}_i\right)^{-1}\left(\sum_{i=1}^{n}\tilde{\ddot{Z}}_i{}'\widetilde{W}_i\right)\right]^{-1}\left(\sum_{i=1}^{n}\widetilde{W}_i{}'\tilde{\ddot{Z}}_i\right)\left(\sum_{i=1}^{n}\tilde{\ddot{Z}}_i{}'\tilde{\ddot{Z}}_i\right)^{-1}\left(\sum_{i=1}^{n}\tilde{\ddot{Z}}_i{}'\,\tilde{Y}_i\right), \qquad (6)$$

where $\tilde{Y}_i = Q_T Y_i$ and analogously for all other variables.

3.2. The FBVR approach

In a similar way to FVR, the FBVR transforms the instrument and then performs a standard TSLS estimation with the transformed instrument. The instrument is transformed in the following way. For all treatment periods ($D_{it} = 1$), FBVR and FVR are equivalent. For all periods prior to the treatment, the values of the instrument are set equal to the value of the instrument in the last untreated period ($T_i - 1$). As in the case of FVR, the instrument values for non-treated units remain unchanged. The data example in Table 2 illustrates this procedure. Thus, the FBVR approach additionally restricts the variation of the instrument backwards. Intuitively, the periods prior to period $T_i - 1$ are considered uninformative. Only the last untreated and first treated values of the instrument are used, as they are the ones that "trigger" the treatment. Thus, the FBVR is closely related to a static approach. Note that the instrument and the treatment are not perfectly correlated. The instrument might have different patterns for different units. It is transformed to follow the (potentially endogenous) pattern of the treatment, but relies on the exogenous values around the implementation of the treatment.

Table 2. FBVR, data example

Period	***Y***	***D***	***Z***	$\boldsymbol{Z_{FBVR}}$
1	65%	0	14,295	13,700
2	64%	0	13,700	13,700
3	70%	1	15,487	15,487
4	72%	1	12,001	15,487

Formally, let the random matrix Q_i have its $T_i - 1$-th column equal to Z_{T_i-1} up to the $T_i - 1$-th row and eventually zero, and its T_i-th column equal to zero up to the $T_i - 1$-th row and then equal to Z_{T_i} (all other columns are equal to zero). Then define $\ddot{\mathcal{L}}_i \coloneqq Q_i \mathcal{L}_i$ for

$T_i \leq T$ and $\dddot{\mathcal{L}}_i \coloneqq \mathcal{L}_i$ for $T_i > T$, and write $\dddot{Z}_i \coloneqq (X_i, \dddot{\mathcal{L}}_i)$. The FBVR estimator is then defined as

$$\hat{\gamma}_{FVR} \coloneqq \left[\left(\sum_{i=1}^{n} \widetilde{W}_i{}'\tilde{\dddot{Z}}_i\right)\left(\sum_{i=1}^{n} \tilde{\dddot{Z}}_i{}'\tilde{\dddot{Z}}_i\right)^{-1}\left(\sum_{i=1}^{n} \tilde{\dddot{Z}}_i{}'\widetilde{W}_i\right)\right]^{-1}\left(\sum_{i=1}^{n} \widetilde{W}_i{}'\tilde{\dddot{Z}}_i\right)\left(\sum_{i=1}^{n} \tilde{\dddot{Z}}_i{}'\tilde{\dddot{Z}}_i\right)^{-1}\left(\sum_{i=1}^{n} \tilde{\dddot{Z}}_i{}'\tilde{Y}_i\right), \tag{7}$$

Remark. There is a subtle but important difference between the FBVR approach and first differencing with instruments. In the latter approach, only observations that exhibit a treatment in the observational period can be used. All other observations are excluded. With the FBVR approach, all observations are used. The point in time of the treatment T_i depends on the vector U_i, as shown below in a discrete choice context. Therefore, excluding observations with $T_i > T$ potentially creates endogeneity. We demonstrate with a simulation that first differencing can in fact produce results that are severely misleading in an empirically relevant example.

3.3. Identification

We adopt the following assumptions to achieve identification:

R1 $\mathbb{E}[U_i | Z_{i1}, Z_{i2}, \dots Z_{iT}, C_i, T_i = k] = 0$ for $t = 1, \dots, T$ and $k = 1, \dots, T$.

R2 a) $rank\left(\mathbb{E}\left[\widetilde{\ddot{Z}}'{}_i\tilde{\ddot{Z}}_i\right]\right) = rank(\sum_{t=1}^{T} \mathbb{E}[\widetilde{\ddot{Z}}'{}_{it}\,\tilde{\ddot{Z}}_{it}]) = L.$

R2 b) $rank\left(\mathbb{E}\left[\widetilde{\dddot{Z}}'{}_i\tilde{\dddot{Z}}_i\right]\right) = rank(\sum_{t=1}^{T} \mathbb{E}[\tilde{\dddot{Z}}'{}_{it}\,\tilde{\dddot{Z}}_{it}]) = L.$

R3 a) $rank\left(\mathbb{E}\left[\widetilde{\ddot{Z}}'{}_i\tilde{X}_i\right]\right) = rank(\sum_{t=1}^{T} \mathbb{E}[\widetilde{\ddot{Z}}'{}_{it}\,\tilde{X}_{it}]) = L.$

R3 b) $rank\left(\mathbb{E}\left[\widetilde{\dddot{Z}}'{}_i\tilde{X}_i\right]\right) = rank(\sum_{t=1}^{T} \mathbb{E}[\tilde{\dddot{Z}}'{}_{it}\,\tilde{X}_{it}]) = L.$

These assumptions are very similar to those used in the standard TSLS approach. This is not surprising as both the FVR and the FBVR approaches rely on the TSLS estimator.

However, we note two important differences. First, as both the FVR and FBVR approaches rely on the potentially endogenous time of treatment T_i, a stronger version of the exclusion restriction is needed. Assumption R1 is stronger than the strict exogeneity assumption 1 stated in Section 2. In particular, R1 implies assumption 1. It requires that the extra information in the period the unit was first treated does not lead to a violation of the exogeneity of the instrument. It precludes the possibility that, conditional on T_i, U_{ik} and Z_{ip} are correlated. The need for a stronger condition arises due to the potentially endogenous adjustment of the instrument.

It is difficult to find lower level conditions that imply R1, because the relations between T_i and the elements of U_i and Z_i are highly nonlinear. Nevertheless, we provide extensive direct and indirect simulation evidence that the assumption is not violated in a variety of empirically relevant cases (see section 4.3). In a setting where the first stage is a discrete choice, we test assumption R1 under different degrees of endogeneity and instrument strength. The data generation process is adopted from established studies and is related to theoretical economic models.

Second, the rank condition R3 is weaker than assumption 3. Thus, there is a trade-off between the strength of the instrument and the validity of the exclusion restriction.

Proposition. *Suppose that either assumptions R1, R2 a), R3 a) or R1, R2 b), R3 b) hold. Then γ is identified.*

The proof is provided in Appendix A.

3.4. Asymptotic properties

The asymptotic properties of the FVR and FBVR approaches can be derived analogously to the TSLS approach.

Asymptotic variance. For expositional simplicity, we assume homoscedasticity of the error term, $\mathbb{E}[U_i U'_i | Z_i, C_i] = \sigma_u^2 I_T$, where σ_u^2 is a positive constant and I_T is the $T \times T$-identity matrix. In addition, assume that W_{it} and Z_{it} are one-dimensional (no covariates other than the treatment). Under the respective set of assumptions, the asymptotic variances of $\sqrt{n}(\hat{\gamma}_{TSLS} - \gamma)$, $\sqrt{n}(\hat{\gamma}_{FVR} - \gamma)$ and $\sqrt{n}(\hat{\gamma}_{FBVR} - \gamma)$ are $\sigma_u^2\left(\mathbb{E}[\widetilde{W}'_i\tilde{Z}_i]\mathbb{E}[\tilde{Z}'_i\tilde{Z}_i]^{-1}\mathbb{E}[\tilde{Z}'_i\widetilde{W}_i]\right)^{-1}$, $\sigma_u^2\left(\mathbb{E}\left[\widetilde{W}'_i\tilde{\tilde{Z}}_i\right]\mathbb{E}[\tilde{\tilde{Z}}'_i\tilde{\tilde{Z}}_i]^{-1}\mathbb{E}[\tilde{\tilde{Z}}'_i\widetilde{W}_i]\right)^{-1}$ and $\sigma_u^2\left(\mathbb{E}\left[\widetilde{W}'_i\tilde{\tilde{\tilde{Z}}}_i\right]\mathbb{E}[\tilde{\tilde{\tilde{Z}}}'_i\tilde{\tilde{\tilde{Z}}}_i]^{-1}\mathbb{E}[\tilde{\tilde{\tilde{Z}}}'_i\widetilde{W}_i]\right)^{-1}$, respectively. Due to the persistence of the treatment, $\mathbb{E}[\widetilde{W}'_i\tilde{Z}_i]$ is much closer to zero in absolute value than its counterparts $\mathbb{E}[\widetilde{W}'_i\tilde{\tilde{Z}}_i]$ and $\mathbb{E}[\widetilde{W}'_i\tilde{\tilde{\tilde{Z}}}_i]$. Thus, the asymptotic variance of the standard TSLS approach should be much larger. In addition, the asymptotic variance of the FBVR approach should be smaller than that of the FVR approach, as the instrument is stronger in the former approach. Since the instrument values are equal for the non-treated units in all three approaches, the gains of the restrictions depend on the number of treated units. These theoretical predictions are confirmed by the simulation results in the next section.

Remark: the FBVR approach imposes a restriction on the maximum number of instruments that can be used. To see this, suppose we have three instruments, $Z_i = \left(Z_{i,1} \dots, Z_{i,T}\right)'$, $V_i = (V_{i,1} \dots, V_{i,T})'$ and $W_i = (W_{i,1} \dots, W_{i,T})'$. Let β_2 be an arbitrary constant and define

$$\beta_1 = \frac{V_{i,T_i}W_{i,T_i-1} - V_{i,T_i-1}W_{i,T_i}}{Z_{i,T_i-1}W_{i,T_i} - Z_{i,T_i}W_{i,T_i-1}}$$

and

$$\beta_3 = \frac{\beta_1 Z_{i,T_i-1} + \beta_2 V_{i,T_i-1}}{W_{i,T_i-1}}.$$

It holds for the transformed instruments

$$\beta_1 \dddot{Z}_i + \beta_2 \dddot{V}_i + \beta_3 \dddot{W}_i = 0.$$

Although β_1 and β_3 are random variables, for each realization of (Z_i, V_i, W_i) there is a linear dependence between the three instruments. In other words, cases with more than two instruments involve a data-induced perfect multicolinearity problem. Note that this is not a real drawback to the FBVR approach, since multiple instruments are only used in IV-models to enhance their strength and the FBVR achieves high strength in a different way.

Consistency and Asymptotic normality. Both the FVR and the FBVR estimators are consistent under assumptions R1-R3. In addition, under weak moment conditions, $\sqrt{n}(\hat{\gamma}_{FVR} - \gamma)$ and $\sqrt{n}(\hat{\gamma}_{FBVR} - \gamma)$ are asymptotically normally distributed. The proofs follow exactly the same steps as for the TSLS estimator and are therefore omitted.

4. Monte Carlo Simulations

In this section, we perform Monte Carlo Simulations to investigate the small sample properties of the FVR and FBVR estimators.

4.1. Data Generating Process

Our data generating process builds on previous work estimating dynamic binary choice models with unobserved heterogeneity (Honoré, 2006; Carro, 2007). It consists of a structural model (8) and a discrete choice model (9):

$$Y_{it} = \alpha D_{it} + \beta X_{it} + C_i + U_{it} \tag{8}$$

$$D_{it} = 1[\mu + \delta D_{it-1} + \gamma \mathcal{L}_{it} + \rho U_{it} + \lambda V_{it} > 0] \tag{9}$$

The sole exogenous regressor is $X_{it} = 5G_{it}$ where G_{it} is an $i.i.d.$ random variable drawn from a continuous uniform distribution $U(0,1)$. C_i equals $\frac{1}{T}\sum_t X_{it}$ to ensure that unobserved heterogeneity is correlated with X_{it}. The error term of the structural equation is $U_{it} \sim i.i.d.N(0,1)$.

The first term of the discrete choice equation μ is the intercept. Other things equal, its value influences the share of units that gets treated. The lagged treatment variable D_{it-1} is included to create persistence in the treatment. Setting δ sufficiently high ensures that treated units remain treated for all subsequent (observed) periods. A strictly exogenous variable $\mathcal{L}_{it} \sim i.i.d.N(0,1)$ may be used as an instrumental variable for the endogenous treatment D_{it}. The strength of $\mathcal{L}_{it}$ is controlled by the size of the parameter γ. The error term of the structural equation U_{it} is included to ensure that D_{it} is endogenous. The higher the value of ρ, the higher the level of endogeneity. Finally, $V_{it} \sim i.i.d.N(0,1)$ is a random noise.

The α and the β parameters are set equal to 1. In the next section, we compute the percentage empirical bias and the 95% confidence interval for the different estimators at different values of γ and ρ. To ensure that the level of endogeneity (instrumental strength) does not change when the value of γ (ρ) is modified, the parameter λ is set equal to $\sqrt{1-\gamma^2-\rho^2}$. This allows us to keep $Var[\gamma \mathcal{L}_{it} + \rho U_{it} + \lambda V_{it}]$ equal to 1 when the value of γ or ρ changes. In our baseline simulation, we generate data for 100 units ($N = 100$) over 15 time periods ($T = 15$) in line with our application dataset. We choose the value for μ so that approximately half of the units receive treatment during the observation period.

4.2. Simulation Results

We draw 1000 samples for each of twelve different empirical settings. Table 3 shows the percentage empirical bias and the 95% confidence interval of the simulated α coefficient estimated by five different fixed-effect estimators: OLS, TSLS, TSLS-probit, FVR, and FBVR.[3] The level of endogeneity ρ equals 0.4 in panel A of Table 3 and 0.6 in panel B. The strength of the instrumental variable, captured by γ, varies within each panel.

Before looking at the relative performance of the four IV-estimators, we perform some consistency checks. The bias of the OLS estimate $\hat{\alpha}_{OLS}$ is higher in panel B (26%) than in panel A (18%). This is consistent with the fact that we set a lower level of endogeneity in panel A. As expected, the distribution of $\hat{\alpha}_{OLS}$ does not change as the strength of the instrument varies. As the level of endogeneity decreases and the strength of the instrument increases, the percentage bias of the four IV-estimators declines and their distributions tighten around the true value of α.

In almost every empirical setting, the percentage bias of $\hat{\alpha}_{FVR}$ and $\hat{\alpha}_{FBVR}$ is lower than the percentage bias of $\hat{\alpha}_{TSLS}$ and $\hat{\alpha}_{TSLS-probit}$. In every setting, $\hat{\alpha}_{FVR}$ and $\hat{\alpha}_{FBVR}$ are clearly more efficient than $\hat{\alpha}_{TSLS}$ and $\hat{\alpha}_{TSLS-probit}$. In addition, Table 3 shows that $\hat{\alpha}_{FBVR}$ performs slightly better than $\hat{\alpha}_{FVR}$ in terms of bias while both estimators are equivalently efficient. All estimators are relatively inefficient when the instrumental variable is relatively weak (γ = 0.1). However, when the instrumental variable is somewhat stronger (γ = 0.3) the 95% confidence interval of TSLS is 30 to 70 times larger than the 95% confidence interval of the FVR and FVBR. This difference decreases as the strength of the instrument increases but remains high even when the instrument is relatively strong. For instance, the 95% confidence

[3] TSLS-probit uses the predictions of a probit model as an instrument for the treatment in a TSLS. See Wooldridge (2002, pp. 623-625) for further details.

interval of TSLS is 6 to 10 times larger than the 95% confidence interval of the FVR and FVBR when γ = 0.6.

In summary, we draw three conclusions based on our Monte Carlo Simulation results. First, when an endogenous treatment is highly persistent, standard approaches such as TSLS and TSLS-probit give uninformative confidence intervals, and this is consistent with theoretical predictions. Second, the FVR and the FBVR estimators prove to be substantially more efficient in a wide range of empirical settings. Third, FVR and FBVR generate estimates that are substantially less biased than OLS given that Z is sufficiently related to D.

We perform several robustness checks to verify the sensitivity of our results. First, we compare the results in Table 3 with a first differencing approach. These results are provided in Appendix B and it is clear that the first differencing approach gives inferior results. When observations are endogenously excluded in a first-difference approach, the bias is 2 to 30 times greater than when using the FBVR. Second, we allow the instrumental variable $\mathcal{L}_{it}$ to be drawn from a non-normal distribution. Third, we generate samples where the share of units treated varies from 30% to 90%. Fourth, we generate samples with different numbers of periods, ranging from T=4 to T=24. When varying the distribution of the instrument, share of units treated and number of time periods, the results are very similar to those presented in Table 3. Details of how these robustness tests were performed, as well as the results, are presented in Appendix C.

Table 3. Simulation results for different values of γ and ρ.

Estimator	Panel A: ρ = 0.4			Panel B: ρ = 0.6		
	% bias	LB	UB	% bias	LB	UB
Empirical setting: $\gamma = 0.1$						
OLS	17.5	1.0	1.3	26.2	1.1	1.4
TSLS	347.5	-529.0	538.0	8.9	-52.4	54.5
TSLS-probit	81.4	-193.8	194.1	40.9	-58.6	61.4
FVR	314.0	-319.0	314.7	18.2	-17.4	19.0
FBVR	68.7	-32.2	35.5	13.5	-20.3	22.0
Empirical setting: $\gamma = 0.2$						
OLS	17.5	1.0	1.3	26.1	1.1	1.4
TSLS	179.9	-75.1	73.5	91.8	-48.6	52.4
TSLS-probit	1029.4	-661.3	683.9	132.5	-79.5	84.1
FVR	0.5	-9.7	11.7	31.1	-8.4	9.8
FBVR	50.5	-34.1	35.1	11.3	-5.9	7.7
Empirical setting: $\gamma = 0.3$						
OLS	17.4	1.0	1.3	26.2	1.1	1.4
TSLS	44.5	-32.9	35.8	9.9	-61.5	63.7
TSLS-probit	105.6	-46.5	50.6	180.4	-144.3	150.0
FVR	11.2	0.0	1.8	15.6	0.0	1.7
FBVR	3.4	-0.2	2.1	8.3	-0.2	2.1
Empirical setting: $\gamma = 0.4$						
OLS	17.6	1.0	1.3	26.3	1.1	1.4
TSLS	7.2	-24.6	26.7	53.0	-38.3	41.4
TSLS-probit	62.0	-44.9	45.6	17.1	-24.1	26.5
FVR	9.5	0.3	1.6	14.2	0.2	1.5
FBVR	3.0	0.3	1.6	4.8	0.3	1.6
Empirical setting: $\gamma = 0.5$						
OLS	17.5	1.0	1.3	26.4	1.1	1.4
TSLS	60.2	-24.0	24.8	19.6	-5.4	7.0
TSLS-probit	78.2	-35.9	36.3	19.4	-4.8	6.4
FVR	9.3	0.4	1.4	14.0	0.4	1.4
FBVR	2.8	0.5	1.4	4.4	0.5	1.4
Empirical setting: $\gamma = 0.6$						
OLS	17.6	1.0	1.3	26.4	1.1	1.4
TSLS	6.0	-2.0	3.9	8.1	-3.0	4.9
TSLS-probit	6.0	-1.9	3.8	7.6	-3.5	5.4
FVR	9.3	0.4	1.4	13.8	0.4	1.3
FBVR	2.6	0.6	1.4	3.9	0.6	1.4

Notes.

(i) $N = 100, T = 15$, 1000 Monte Carlo replications.

(ii) LB denotes 95% lower bound and UB denotes 95% upper bound.

(iii) $\mathcal{L}_{it} \sim i.i.d. N(0,1)$.

(iv) $\mu = -1.662$ to obtain 50% of unit treated.

(v) $\alpha = 1$.

4.3. Empirical test of the identifying assumption

In section 3.3, we show that assumption R1 is crucial for identification and that it guarantees consistent estimation when using the FVR and FBVR approaches. Similar to the standard strict exogeneity assumption, assumption R1 is non-testable in an empirical context since the error term is unobserved. In this section, we provide evidence that R1 holds under (8) and (9).

Note that R1 implies the following assumption:

R1' $\mathbb{E}\left[Z_{it}U_{ip} \mid T_i = k\right] = 0$ for $t = 1, \dots, T; p = 1, \dots, T$ and $k = 1, \dots, T$

We generate 100 samples of 1,000 units for 11 periods. The data generating process is the same as in previous sections. We set $\rho = 0.4$, $\gamma = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k}\sum_{i\in K} Z_{it}U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k-1}\sum_{i\in K}\left(Z_{it}U_{ip} - \bar{P}_{ktp}\right)^2}$ where n_k is the number of units treated in period k. Under R1', the statistic $\theta_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student($n_k - 1$) distribution. Appendix D reports the number of rejections together with sample means and variances of the test statistics for some values of t, k, p. For each combination of t, k, p, we fail to reject R1'. Thus, in this setting, additional conditioning on the point in time of treatment does not lead to a violation of the exclusion restriction. This is a novel result.

5. Empirical investigation: the effect of utility divestiture on nuclear reactor unavailability

In this section, we evaluate the effect of electricity utility divestiture on nuclear reactors' unavailability factor in the U.S.[4] Such an ex-post evaluation is valuable to reactor stakeholders since it gives information about the effect of asset divestiture on reactor performance. The results can also be used to inform policy-makers about the treatment's welfare and environmental effects.[5] None of the divestiture actions are reversed during the sample period we consider and this treatment is therefore persistent.

Economic theory predicts that divestiture increases competition, which improves economic performance. Green (1996) employs a supply function equilibrium model and finds that partial divestiture leads to a reduction of deadweight loss. Borenstein and Bushnell (1999) model the California electricity market after deregulation as a static Cournot market with a competitive fringe and they find that divestiture can reduce market power. More recently, Zhang (2007) explains that restructured U.S. reactors are no longer able to simply pass on the costs of repair and maintenance performed during outages, and that this has increased incentives to reduce outages.

Two studies have previously investigated the effect of divestiture, and closely related reforms, on nuclear reactor performance in the U.S.. Zhang (2007) investigates how the reactor availability factor is affected by the intended and actual implementation of retail competition and she relaxes the assumption that the deregulatory reform is exogenous. Her results, based on standard TSLS models, indicate that increased retail competition increases

[4] Detailed descriptions of this market transformation process have been presented by several authors, e.g. Delmas and Tokat (2005), Zhang (2007), Davis and Wolfram (2012), and references therein.

[5] See Davis and Wolfram (2012) for a quantification of the effects of nuclear reactor divestiture on electricity prices and CO_2 emissions.

reactor availability but the coefficient of the endogenous reform variable is never significant. Davis and Wolfram (2012) focus solely on utility divestiture and they assume that divestiture is exogenous. Their OLS results suggest that divestiture increases the availability factor by 10 percentage points, and the effect is statistically significant.[6] Thus, these two studies illustrate two empirical traps: first that the instrument is weak, and second that the endogeneity problem is ignored.

5.1. Data

We use a balanced sample with annual data that represents all U.S. nuclear reactors from 1994 to 2011. The first utility to divest its assets did so in 1999, and the last one divested in 2007. Of a total 103 U.S. nuclear reactors, 47% were subject to divestiture during this time period. Data is collected from different sources. We use annual data about nuclear reactors' outage duration, location and technical characteristics from the IAEA PRIS database. This includes the state where the reactors are located, the year they were first connected to the grid, and technical characteristics in terms of technology (PWR versus BWR), containment structure, and steam generator type. Data on the year of divestiture is collected from Davis and Wolfram (2012).[7] Finally, data about state level political majority comes from the US census bureau.

Table 4 presents descriptive statistics and information about relevant variables. The maximum value of *UF* is 100, indicating no production during a whole year. A closer examination of the data reveals 20 such observations. In the subsequent estimations we either

[6] Studies have also been made on the effect of divestiture in other industries. Soetevent et al. (2014) analyze the impact of divestiture on Dutch highway gasoline stations and find that divestiture lowers the price of divested stations and neighboring stations.
[7] These data have been cross-checked using the Nuclear Energy Institute website for divestiture.

include a dummy variable to control for these observations or exclude them as a test of robustness. It should also be emphasized that as the minimum value for Age is 1, no reactor has entered the market during the sample period.

Table 4. Descriptive statistics

Variable	Description	Level of aggregation	No. obs.	Mean	Std. Dev.	Min	Max
UF	100-(Actual operating hours / Potential operating hours)×100	Reactor	1851	12.397	15.132	0	100
$Divest$	Equals 1 if the reactor is divested. 0 otherwise	Reactor	2369	0.204	0.403	0	1
Age	Age of the reactor (in years)	Reactor	1851	23.733	8.299	1	43
Ind	Share of state level electricity consumption from industrial consumers	State	2266	21.770	7.707	4.974	48.250
Rep	Equals 1 if both state senate and house of representatives have a Republican majority	State	2266	0.299	0.458	0	1

5.2. Model and main results

Our structural equation uses the reactor unavailability factor (*UF*) as dependent variable and utility divestiture (*Divest*) as an independent variable. We control for reactor age and include both Age and Age^2. This is because a newly built reactor may have to be calibrated to site-specific conditions at the beginning of its life. After the calibration period, the probability of disruption declines. As the reactor gets older, disruptions may increase again due to greater demand for repairs and maintenance. A further control (*S*) is an indicator that takes the value 1 when *UF*=100, i.e. when reactors have not produced any electricity during a given year.

Moreover, we include year fixed effects $\boldsymbol{\eta}_t$ and reactor fixed effects λ_i. Year fixed effects capture all regulatory and economic variations at federal level. Reactor fixed effects capture stable conditions, such as technology choices, firm culture and geographical characteristics. This specification is flexible in the sense that it allows analysts to include regressors that are correlated with λ_i, such as maintenance costs/procedures that are influenced by technology, but that are unobserved in our data.

The equation of interest can then be written as:

$$UF_{it} = \beta_0 1[Divest_{it}] + \beta_1 Age_{it} + \beta_2 Age_{it}^2 + \beta_3 S_{it} + \boldsymbol{\eta}_t \gamma_t + \lambda_i + \varepsilon_{it}, \quad (10)$$

where *i* is the reactor, *t* is the year and ε_{it} are the random errors. In this model, $Divest_{it}$ is potentially correlated with ε_{it}. This is because $Divest_{it}$ is a function of the state-level electricity price in year *t* (Ando and Palmer, 1998; Delmas and Tokat, 2005; Fabrizio et al., 2007; Damsgaard, 2003), but state-level electricity price in year *t* is also a function of UF_{it} (Zhang, 2007). The reason why a high electricity price increases the likelihood of divestiture is that it tends to be interpreted as a sign of market failure that triggers policy action. The positive impact of UF_{it} on the electricity price in year *t* is explained by the fact that nuclear power is a baseload component in the electricity generation mix. More expensive sources of energy have to be used whenever reactor operations are disrupted. A similar type of simultaneity applies to the nuclear sector's lobby group activity: more intense lobby activity in year *t* reduces regulatory pressure on the industry and, thus, increases UF_{it}. At the same time, increased lobby group activity reduces the likelihood of divestiture. Since both state-level electricity and lobby group activity are unobserved in our data set, $Divest_{it}$ becomes endogenous.

In our base specification, we take the share of state level electricity consumption by industrial customers in the previous period (Ind_{t-1}) as our only instrumental variable. The choice of instrument is based on the political economy of the electricity market restructuring process. In particular, Joskow (1997) stresses the importance of interest groups that supported electricity market reforms in the U.S. during the 1990s. At this time, the expectation was that large industrial consumers would benefit from stronger competition and thus more actively support electricity market restructuring. This instrument is also used by Zhang (2007).

The estimations of (10) using the FVR and FBVR approaches are compared to the following alternative approaches: 1) OLS, which is used by Davis and Wolfram (2012), i.e. where endogeneity is ignored, 2) TSLS, which is used by Zhang (2007), i.e. where treatment persistence is ignored, and 3) TSLS-probit, which also ignores treatment persistence, but is potentially more efficient than the TSLS.

The main results are presented in Table 5. If we assume that FBVR provides the least biased estimate, then the OLS estimate appears to be slightly upward-biased. This may be because the nuclear industry's lobby group activity is unobserved and is negatively related to $Divest_{it}$ and positively related to UF_{it}. The second noteworthy observation is that the $SE(\beta_0)$ for the TSLS is about twice as large as for the TSLS-probit and FVR approaches, and $SE(\beta_0)$ is about six times as large as for the FBVR approach. The SEs of the TSLS, TSLS-probit and the FVR approaches are so large we cannot statistically distinguish the divestiture effect from zero. The conclusion based on the FBVR is that the divestiture of electricity utilities reduces the unavailability factor of the nuclear reactors by 7.6%.

Table 5. Estimation output of model (10)

Variable	OLS	TSLS	TSLS-probit	FVR	FBVR
$Divest_t$	-7.143*** (1.452)	7.095 (13.460)	5.762 (6.976)	-6.771 (5.594)	-7.624*** (2.202)
Age_t	0.150 (0.218)	-0.216 (0.419)	-0.182 (0.243)	0.141 (0.252)	0.162 (0.210)
Age_t^2	-0.010** (0.005)	-0.010** (0.004)	-0.010** (0.004)	-0.010** (0.004)	-0.010** (0.004)
Year dummies	Yes	Yes	Yes	Yes	Yes
Treatment of obs. where UF=100	Dum. Var.	Dum. Var.	Dum. Var.	Dum. Var.	Dum. Var.
R^2	0.39	0.37	0.38	0.44	0.44
No. obs.	1851	1851	1851	1851	1851

Notes: Dependent variable is *UF*. *UF* represents total number of outage hours divided by maximum potential generation hours. * p < 0.10, ** p < 0.05, *** p < 0.01. SE in brackets are robust to heteroskedasticity and autocorrelation with a Bartlett bandwidth = 2.

Finally, we check the robustness of our empirical findings. One first robustness test is to use an alternative instrumental variable. Thus, we replace Ind_{t-1} with a dummy variable that indicates whether the state has a Republican majority (Rep_{t-1}). The results are similar to our base estimation. The coefficient estimated by FBVR, which equals -11, is lower but not statistically different from the coefficient obtained in Table 5. The methodology and detailed results are given in Appendix E.

The results presented so far rely on the assumption that untreated units are completely unaffected by reactors that are treated. As an additional test of robustness, we evaluate the reasonableness of this assumption. A detailed description of the methodology and results is given in Appendix F. The results indicate that we cannot reject the null hypothesis that there are no spillover effects from the divestiture of other reactors.

6. Conclusions

Policies are often endogenous and persistent. This leads to a weak instrumental variable problem when the values of available instrument(s) change from period to period. In this paper, we develop two approaches to strengthen the instrument in this context by removing unnecessary instrument variation. In the FVR approach, we first set the values of the instrument in all treated periods equal to the value of the instrument in the first treated period. Next, TSLS is performed with the transformed instrument. In the FBVR approach, the instrument is also transformed backward by taking the instrumental value in the last untreated period and copy it to all previous periods.

We theoretically prove identification and derive asymptotic properties. The main intuition of the approaches does not depend on linearity, suggesting that similar techniques can be used for a variety of models. Moreover, our approaches could be used to evaluate structural models.

We also evaluate small sample properties for the FVR/FBVR approaches through Monte Carlo simulations. FVR/FBVR empirical standard errors are up to 70 times smaller than for TSLS, and FVR/FBVR empirical bias is up to 10 times smaller than for OLS. These results are largely robust to the instrument distribution, endogeneity level and instrument strength.

We use the FVR/FBVR approaches to evaluate the effect of the divestiture of nuclear reactors in the U.S. implemented in the 1990s and 2000s. Studies that have previously evaluated this policy reform have either ignored treatment endogeneity or produced uninformative confidence intervals. We find that divestiture has reduced the reactor unavailability factor by approximately 7.6% and the effect is statistically significant.

7. References

Akay, A., (2012), Finite-Sample comparison of alternative methods for estimating dynamic panel data models, Journal of Applied Econometrics, Vol. 27(7), pp. 1189?1204.

Anderson, T.W. and Rubin, H., (1949), Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, The Annals of Mathematical Statistics, Vol. 20(1), pp. 46-63.

Ando, A.W. and Palmer, K.L., (1998), Getting on the map: the political economy of state-level electricity restructuring, Discussion Paper 98-19-REV, Resources for the future: Washington.

Angrist, J.D. and Krueger, A.B., (2001), Instrumental variables and the search for identification: from supply and demand to natural experiments, Journal of Economic Perspectives, Vol. 15(4), pp. 69-85.

Bekker, P.A., (1994), Alternative approximations to the distributions of instrumental variable estimators, Econometrica, Vol. 62(3), pp. 657-681.

Bertrand, M., (2004), From the invisible handshake to the invisible hand? How import competition changes the employment relationship, Journal of Labour Economics, Vol. 22(4), pp. 723-765.

Borenstein, S. and Bushnell, J. (1999). An Empirical Analysis of the Potential for Market Power in California's Electricity Industry. The Journal of Industrial Economics, 47 (3): 285-323

Bound, J., Jaeger, D.A. and Baker, R.M., (1995), Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak, Journal of the American Statistical Association, Vol. 90(430), pp. 443-450.

Carro, J. M. (2007), Estimating dynamic panel data discrete choice models with fixed effects, Journal of Econometrics, 140 (2): 503-528.

Coate, S. and Morris, S., (1999), Policy persistence, American Economic Review, Vol. 89, pp. 1327-1336.

Craig and Savage (2013), Market restructuring, competition and the efficiency of electricity generation: Plant-level evidence from the United States 1996 to 2006, The Energy Journal, 34 (1): 1-30.

Damsgaard, N., (2003), Why deregulation? Economics and politics in retail electricity markets, mimeo Stockholm School of Economics.

Davis, L. and Wolfram, C. (2012), Deregulation, consolidation, and efficiency: Evidence from US nuclear power, American Economic Journal: Applied Economics, 4 (4): 194–225.

Delmas, M. and Tokat, Y., (2005), Deregulation, governance structures, and efficiency: the electric utility sector, Strategic Management journal, Vol. 26, pp. 441-460.

Fabrizio, K.R., Rose,N.L and Wolfram C.D. (2007), Do markets reduce costs? Assessing the impact of regulatory restructuring on us electric generation efficiency, American Economic Review, 97 (4): 1250-1277.

Fernandez, R. and Rodrik, D., (1991), Policy persistence and rent extraction: status quo bias in the persistence of individual-specific uncertainty. American Economic Review, Vol 81, pp. 1146-1155.

Flores-Lagunes, A., (2007), Finite sample evidence of IV estimators under weak instruments, Journal of Applied Econometrics, Vol. 22(3), pp. 677–694.

Fredriksson, P. and Johansson, P., (2008), Dynamic Treatment Assignment. Journal of Business & Economic Statistics, 26, pp. 435–445.

Green, R. (1996), Increasing Competition in the British Electricity Spot Market. The Journal of Industrial Economics, 44 (2): 205-216

Hahn, J. and Hausman, J., (2003), Weak Instruments: Diagnosis and Cures in Empirical Econometrics, The American Economic Review, Papers and Proceedings, Vol. 93(2), pp. 118-125.

Heckman, J.J. and Navarro, S., (2007), Dynamic discrete choice and dynamic treatment effects, Journal of Econometrics, Vol. 136, pp. 341-396.

Honoré, B. E. and Tamer, E., (2006), Bounds on parameters in panel dynamic discrete choice models, Econometrica, 74 (3): 611-629

Honore, B.E. and Kyriazidou, E., (2000), Panel data discrete choice models with lagged dependent variables, Econometrica, Vol. 74, pp. 611-629.

Joskow, P.L. (1997), Restructuring, competition and regulatory reform in the U.S. electricity sector,Journal of Economic Perspectives, 11 (3): 119–138.

Raphael, S. and Winter-Ebmer, R., (2001), Identifying the Effect of Unemployment on Crime, Journal of Law and Economics, Vol. 44(1), pp. 259-283.

Ratkovic, M. and Shiraito, Y., (2014), Strengthening weak instruments by modelling compliance, mimeo, Princeton University.

Revenga, A., (1990), Essays on labor market adjustment and open economics. PhD dissertation Harvard University, Economics Department.

Revenga, A., (1992), Exporting jobs? The impact of import competition on employment and wages in U.S. manufacturing, Quarterly Journal of Economics, Vol. 107(1), pp. 255-284.

Sianesi, B. (2004). An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. The Review of Economics and Statistics, 86(1), pp. 133–155.

Soetevent, A.R., Haan, M.A. and Heijnen, P. (2014). Do Auctions and Forced Divestitures Increase Competition? Evidence for Retail Gasoline Markets. The Journal of Industrial Economics, 62 (3): 467-502

Staiger, D. and Stock, J.H., (1997), Instrumental variable regression with weak instruments, Econometrica, Vol. 65(3), pp. 557-586.

Stock, J.H., Wright, J.H. and Yogo, M., (2002), A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, Journal of Business & Economic Statistics, Vol. 20(4), pp. 518-529.

Taber, C. R., (2000), Semiparametric identification and heterogeneity in discrete choice dynamic programming models, Journal of Econometrics, 96(2), pp. 201–229.

Zhang, F. (2007), Does electricity restructuring work? Evidence from the US nuclear energy industry, Journal of Industrial Economics, 55 (3): 397-418.

Warren, P.L. and Wilkening, T.S., (2012), Regulatory fog: the role of information in regulatory persistence, Journal of Economic Behavior & Organization, Vol. 84, pp. 840-856.

Wooldridge. J.M.. 2002. Econometric analysis of cross section and panel data. Cambridge. MA: MIT Press.

8. Appendices

Appendix A. Proof of Proposition in Section 3.3.

We provide proof for the case of a single (endogenous) covariate and a single instrument. The generalization to the case of additional covariates and multiple instruments follows the same steps and is omitted. Assume first that $\bar{T} = T$. It holds

$$\mathbb{E}\left[1\{T_i = k\}Z_{il}U_{ip}\right] = 0 \tag{A.1}$$

for all $k, l, p \in \{1,2, \dots, T\}$. This follows from

$$\mathbb{E}\left[1\{T_i = k\}Z_{il}U_{ip}\right] = \mathbb{E}\left[\mathbb{E}\left[1\left\{\{T_i = k\}Z_{il}U_{ip}\middle|T_i\right]\right] = \sum_{t=1}^{T}\mathbb{E}\left[1\{t = k\}\,Z_{il}U_{ip}\middle|T_i = t\right]P\{T_i = t\} = 0 + \mathbb{E}\left[Z_{il}U_{ip}\middle|T_i = k\right]P\{T_i = k\} = 0.$$

The last equality holds due to assumption R1 and

$$\mathbb{E}\left[Z_{il}U_{ip}\middle|T_i = k\right] = \mathbb{E}[\mathbb{E}[Z_{il}U_{ip}|Z_{il}, T_i = k]|T_i = k] = \mathbb{E}[Z_{il}\mathbb{E}[U_{ip}|Z_{il}, T_i = k]|T_i = k].$$

Due to analogical arguments, it holds that

$$\mathbb{E}\left[T_iZ_{il}U_{ip}\right] = 0 \tag{A.2}$$

for all $l, p \in \{1, 2, \dots, T\}$.

Identification of FVR: Multiply model (4) by $\tilde{\tilde{Z}}_i$ to obtain

$$\tilde{\tilde{Z}}'_i\tilde{Y}_i = \tilde{\tilde{Z}}'_i\tilde{W}_i\gamma + \tilde{\tilde{Z}}'_i\tilde{U}_i. \tag{A.3}$$

The identification proof follows the same steps as the proof for the standard TSLS model. We now show that $\mathbb{E}\left[\tilde{\tilde{Z}}'_i\tilde{U}_i\right] = 0$. The k-th element of the vector $\tilde{\tilde{Z}}_i$ is equal to

$$Z_{ik}1\{T_i \geq k\} + Z_{iT_i}1\{T_i < k\} - T^{-1}\sum_{l=1}^{T_i}Z_{il} - T^{-1}(T - T_i)Z_{iT_i}. \tag{A.4}$$

We now prove that the expectation of the product of any of these four terms with U_{ip} is equal to zero for any p. For the first term, it holds because of (A.3) and because $1\{T_i \geq k\} = \sum_{l=k}^{T}1\{T_i = l\}$. For the second term, observe that $Z_{iT_i} = \sum_{l=1}^{T}1\{T_i = l\}Z_{il}$ and $1\{T_i = l\}1\{T_i = m\} = 0$ whenever $l \neq m$ and then analogous argument as for the first term applies.

For the third term, observe that $\sum_{l=1}^{T_i} Z_{il} = \sum_{k=1}^{T} 1\{T_i = k\} \sum_{l=1}^{k} Z_{il}$ and hence $\mathbb{E}\left[\sum_{l=1}^{T_i} Z_{il}\, U_{ip}\right] = \sum_{k=1}^{T} \sum_{l=1}^{k} \mathbb{E}\left[1\{T_i = k\} Z_{il} U_{ip}\right] = 0$. Finally, $\mathbb{E}\left[Z_{il} U_{ip}\right] = 0$ and with (A.4) we obtain $\mathbb{E}\left[T^{-1}(T - T_i) Z_{iT_i} Z_{il} U_{ip}\right] = 0$. Therefore, all summands in the sum $\tilde{\tilde{Z}}'_i \tilde{U}_i$ have an expectation of zero. Thus $\mathbb{E}\left[\tilde{\tilde{Z}}'_i \tilde{U}_i\right] = 0$. Using assumptions R2 a) and R3 a), we finally obtain

$$\gamma = \mathbb{E}[\Pi \tilde{\tilde{Z}}'_i \widetilde{W}_i]^{-1} \mathbb{E}[\Pi \tilde{\tilde{Z}}'_i \tilde{Y}_i], \qquad \text{(A.5)}$$

with $\Pi \coloneqq \mathbb{E}[\tilde{\tilde{Z}}'_i \tilde{\tilde{Z}}_i]^{-1} \mathbb{E}[\tilde{\tilde{Z}}'_i \widetilde{W}_i]$.

Identification of FBVR: The proof follows exactly the same steps as the FVR approach. Observe that the k-th element of $\ddot{\tilde{Z}}_i$ is equal to

$$\check{Z}_{ik} - \frac{1}{T}\Big((T_i - 1) Z_{iT_i - 1} + (T - (T_i - 1)) Z_{iT_i}\Big),$$

where $\check{Z}_{ik} = 1\{T_i > k\} Z_{iT_i - 1} + 1\{T_i \leq k\} Z_{iT_i}$. Therefore, showing that $\mathbb{E}\left[\ddot{\tilde{Z}}'_i \tilde{U}_i\right] = 0$ amounts to showing that

$$\mathbb{E}\left[1\{T_i > k\} Z_{iT_i - 1} U_{it}\right] - \mathbb{E}\left[1\{T_i > k\} Z_{iT_i} U_{it}\right] + \mathbb{E}\left[T_i Z_{iT_i - 1} U_{it}\right] - \mathbb{E}\left[T_i Z_{iT_i} U_{it}\right] = 0, \qquad \text{(A.6)}$$

which has been established above.

The arguments are analogous when $\bar{T} > T$. This can be shown by introducing $1\{T_i > T\}$ and $1\{T_i \leq T\}$. In the first case, all expressions are the same as in the standard TSLS case, and the second case is as above.

■

Appendix B. Monte Carlo simulations using first difference approach

Table B.1. Simulation results with different γ and ρ values

	Panel A: ρ = 0.4			**Panel B:** ρ = 0.6		
Estimator	**% bias**	**LB**	**UB**	**% bias**	**LB**	**UB**
Empirical setting: $\gamma = 0.1$						
OLS	87.4	1.8	2.0	130.0	2.2	2.4
TSLS	61.0	0.6	2.6	71.4	0.9	2.6
TSLS-probit	44.8	-71.8	72.9	174.0	-21.0	26.5
FVR	61.0	0.6	2.6	71.4	0.9	2.6
FBVR	61.0	0.6	2.6	71.4	0.9	2.6
Empirical setting: $\gamma = 0.2$						
OLS	87.3	1.8	2.0	130.1	2.2	2.4
TSLS	57.6	1.1	2.0	73.1	1.3	2.2
TSLS-probit	68.9	0.2	3.2	101.4	0.4	3.6
FVR	57.6	1.1	2.0	73.1	1.3	2.2
FBVR	57.6	1.1	2.0	73.1	1.3	2.2
Empirical setting: $\gamma = 0.3$						
OLS	87.7	1.8	2.0	129.8	2.2	2.4
TSLS	54.3	1.2	1.9	73.5	1.4	2.1
TSLS-probit	61.2	0.9	2.3	83.5	1.2	2.5
FVR	54.3	1.2	1.9	73.5	1.4	2.1
FBVR	54.3	1.2	1.9	73.5	1.4	2.1
Empirical setting: $\gamma = 0.4$						
OLS	87.3	1.8	2.0	130.6	2.2	2.4
TSLS	53.7	1.3	1.8	75.4	1.5	2.0
TSLS-probit	53.8	1.1	2.0	78.3	1.3	2.2
FVR	53.7	1.3	1.8	75.4	1.5	2.0
FBVR	53.7	1.3	1.8	75.4	1.5	2.0
Empirical setting: $\gamma = 0.5$						
OLS	87.4	1.8	2.0	130.2	2.2	2.4
TSLS	51.7	1.3	1.7	74.2	1.5	2.0
TSLS-probit	49.7	1.2	1.8	72.2	1.4	2.0
FVR	51.7	1.3	1.7	74.2	1.5	2.0
FBVR	51.7	1.3	1.7	74.2	1.5	2.0
Empirical setting: $\gamma = 0.6$						
OLS	87.3	1.8	2.0	130.2	2.2	2.4
TSLS	50.7	1.3	1.7	72.8	1.5	1.9
TSLS-probit	47.4	1.2	1.7	71.3	1.5	2.0
FVR	51.7	1.3	1.7	72.8	1.5	1.9
FBVR	50.7	1.3	1.7	72.8	1.5	1.9

Notes.

(i) $N = 100, T = 15$. 1000 Monte Carlo replications.
(ii) LB denotes 95% lower bound and UB denotes 95% upper bound.
(iii) $\mathcal{L}_{it} \sim i.i.d. N(0,1)$.
(iv) $\mu = -1.662$ to obtain 50% of unit treated.
(v) $\alpha = 1$.

Appendix C. Monte Carlo Simulations: robustness checks

To check the robustness of our results, we run the simulations under different conditions. Honoré and Kyriazidou (2000) suggest that normally distributed explanatory variables produce smaller bias than non-normally distributed variables. In Table C.1, we report simulations where $\mathcal{L}_{it}$ is drawn from a χ^2-distribution, which is skewed. Like Akay (2012), we standardize this distribution by calculating $\frac{\chi^2_{(1)}-1}{\sqrt{2}}$ to facilitate comparison with the N(0,1) distribution. These results are similar to those shown in Table 3. The only significant difference is that the variances of TSLS and TSLS-probit are smaller than when $\mathcal{L}_{it}$ is drawn from a normal distribution.

In our baseline results, the share of treated units equals 50%. As the share of treated units reduces, the dataset used for F(V)BR and the dataset used for TSLS become more similar. This is because the transformed instrument is applied to a lower proportion of units. In Table C.2, we perform simulations in which the share of treated units varies from 30% to 90%. We start at 30% because estimations with TSLS and TSLS-probit do not converge at lower shares. Table 8 shows that the efficiency of FVR and FBVR does not vary significantly between 50% and 90%. At 30%, the confidence intervals of the FVR and FBVR estimators are slightly wider, whereas the TSLS and TSLS-probit intervals are uninformative in all empirical settings.

As a final robustness test, we run the simulations for different numbers of periods, covering the range from *T*=4 to *T*=24. The results are given in Table C.3. For every empirical setting, FVR and FBVR are substantially more efficient than TSLS and TSLS-probit.

Table C.1. Simulation results with different γ and ρ values, χ^2-distribution

	Panel A: ρ = 0.4			**Panel B:** ρ = 0.6		
Estimator	**% bias**	**LB**	**UB**	**% bias**	**LB**	**UB**
Empirical setting: γ = 0.1						
OLS	17.3	1.0	1.3	26.0	1.1	1.4
TSLS	45.4	-82.2	85.1	1027.1	-604.2	626.7
TSLS-probit	164.5	-214.7	213.4	360.9	-503.6	512.8
FVR	213.5	-98.2	96.0	66.3	-54.6	55.3
FBVR	84.6	-33.5	37.2	4.3	-11.5	13.5
Empirical setting: γ = 0.2						
OLS	16.9	1.0	1.3	25.8	1.1	1.4
TSLS	81.0	-73.7	74.1	70.8	-39.4	40.0
TSLS-probit	76.8	-63.9	67.4	85.4	-50.8	54.5
FVR	8.5	-1.8	3.7	15.2	-1.2	2.9
FBVR	5.7	-0.9	2.8	9.6	-7.3	9.1
Empirical setting: γ = 0.3						
OLS	16.2	1.0	1.3	24.6	1.1	1.4
TSLS	11.8	-5.6	7.3	79.0	-55.8	59.4
TSLS-probit	140.2	-89.1	93.9	1.6	-7.1	9.1
FVR	6.6	0.4	1.5	10.5	0.3	1.5
FBVR	2.3	0.4	1.6	3.6	0.4	1.6
Empirical setting: γ = 0.4						
OLS	15.3	1.0	1.3	23.0	1.1	1.4
TSLS	4.1	-1.4	3.3	9.4	-2.1	4.0
TSLS-probit	4.6	-1.2	3.1	8.6	-1.9	3.7
FVR	4.7	0.6	1.3	7.1	0.5	1.3
FBVR	1.2	0.6	1.4	1.7	0.6	1.4
Empirical setting: γ = 0.5						
OLS	14.1	1.0	1.3	21.2	1.0	1.4
TSLS	2.1	-0.5	2.4	3.6	-0.5	2.4
TSLS-probit	1.8	-0.4	2.4	3.1	-0.5	2.4
FVR	3.8	0.7	1.3	5.8	0.6	1.3
FBVR	0.6	0.7	1.3	0.9	0.7	1.3
Empirical setting: γ = 0.6						
OLS	12.5	1.0	1.3	18.9	1.0	1.4
TSLS	1.3	-0.1	2.1	1.8	-0.2	2.1
TSLS-probit	1.1	-0.1	2.1	1.5	-0.1	2.1
FVR	3.8	0.7	1.3	5.0	0.7	1.2
FBVR	0.3	0.7	1.2	0.5	0.8	1.2

1000 Monte Carlo replications. LB denotes 95% lower bound and UB denotes 95% upper bound. $\mathcal{L}_{it} \sim i.i.d.\ \frac{\chi^2_{(1)}-1}{\sqrt{2}}$. $\alpha = 1$.

Table C.2. Simulation results with different shares of units treated

Estimator	Panel A: ρ = 0.4, γ = 0.4 % bias	LB	UB	Panel B: ρ = 0.6, γ = 0.4 % bias	LB	UB
Empirical setting: 30 % treated						
OLS	18.5	1.0	1.4	31.3	27.4	1.0
TSLS	24.2	-32.0	34.5	815.5	-418.8	404.5
TSLS-probit	189.7	-104.1	102.3	206.5	-55.9	53.7
FVR	13.3	-0.2	1.9	20.7	-0.2	1.8
FBVR	11.5	-0.6	2.3	16.3	-0.4	2.1
Empirical setting: 50 % treated						
OLS	17.6	1.0	1.3	26.4	1.1	1.4
TSLS	6.0	-2.0	3.9	8.1	-3.0	4.9
TSLS-probit	6.0	-1.9	3.8	7.6	-3.5	5.4
FVR	9.3	0.4	1.4	13.8	0.4	1.3
FBVR	2.6	0.6	1.4	3.9	0.6	1.4
Empirical setting: 60 % treated						
OLS	17.1	1.0	1.3	25.8	1.1	1.4
TSLS	39.0	-12.9	14.1	4.8	-13.0	15.1
TSLS-probit	41.1	-13.2	14.4	26.6	-12.8	14.3
FVR	8.7	0.4	1.5	12.7	0.3	1.4
FBVR	0.7	0.5	1.5	1.3	0.5	1.5
Empirical setting: 75 % treated						
OLS	16.7	1.0	1.3	24.8	1.1	1.4
TSLS	9.0	-9.6	11.8	6.6	-6.1	8.0
TSLS-probit	9.5	-9.4	11.6	6.3	-6.1	8.0
FVR	5.9	0.5	1.4	9.0	0.5	1.4
FBVR	3.8	0.6	1.4	5.4	0.7	1.4
Empirical setting: 90 % treated						
OLS	16.6	1.0	1.3	24.8	1.1	1.4
TSLS	88.1	-65.2	69.0	9.9	-2.5	4.3
TSLS-probit	16.4	-6.8	8.5	9.9	-2.4	4.2
FVR	2.2	0.6	1.4	3.5	0.6	1.3
FBVR	9.3	0.8	1.4	13.3	0.8	1.4

1000 Monte Carlo replications. LB denotes 95% lower bound and UB denotes 95% upper bound. $\mathcal{L}_{it} \sim i.i.d.\ N(0,1)$. We change the value of μ to obtain the desired shares of units treated.

Table C.3. Simulation results with different *T*.

	Panel A: ρ = 0.6, γ = 0.6			Panel B: ρ = 0.6, γ = 0.6		
Estimator	**% bias**	**LB**	**UB**	**% bias**	**LB**	**UB**
Empirical setting: *T*=4						
OLS	51.3	0.9	2.1	77.8	1.2	2.3
TSLS	18.0	-34.5	36.1	47.6	-17.6	18.6
TSLS-probit	51.1	-52.6	55.7	63.3	-14.4	15.1
FVR	29.5	-2.3	3.7	46.4	-2.7	3.8
FBVR	34.5	-3.0	4.4	38.5	-9.1	10.3
Empirical setting: *T*=8						
OLS	28.8	1.0	1.6	43.3	1.1	1.7
TSLS	17.4	-5.7	8.1	7.6	-5.5	7.3
TSLS-probit	1920.6	-1160.0	1200.4	1.7	-4.0	6.0
FVR	13.0	-0.1	1.8	22.3	-0.2	1.8
FBVR	8.8	-0.1	1.9	15.0	-0.1	1.8
Empirical setting: *T*=12						
OLS	20.8	1.0	1.4	31.6	1.1	1.5
TSLS	4.4	-2.2	4.1	8.0	-2.1	4.0
TSLS-probit	5.7	-2.3	4.2	9.3	-2.4	4.3
FVR	11.8	0.3	1.4	18.0	0.3	1.4
FBVR	6.8	0.4	1.5	9.4	0.4	1.4
Empirical setting: *T*=16						
OLS	16.5	1.0	1.3	24.1	1.1	1.4
TSLS	0.2	-2.7	4.7	14.0	-6.7	9.0
TSLS-probit	1.8	-3.5	5.5	18.8	-8.7	11.1
FVR	8.4	0.5	1.3	13.6	0.5	1.3
FBVR	2.2	0.6	1.3	4.2	0.6	1.3
Empirical setting: *T*=20						
OLS	13.6	1.0	1.3	20.4	1.1	1.3
TSLS	5.7	-1.9	4.0	7.7	-6.3	8.2
TSLS-probit	6.8	-2.0	4.1	14.7	-12.0	13.7
FVR	7.0	0.6	1.2	10.6	0.6	1.2
FBVR	0.6	0.7	1.3	1.0	0.7	1.3
Empirical setting: *T*=24						
OLS	11.8	1.0	1.2	17.8	1.1	1.3
TSLS	0.5	-6.4	8.4	7.9	-2.2	4.0
TSLS-probit	1.6	-4.6	6.6	7.8	-2.1	4.0
FVR	7.0	0.6	1.2	9.6	0.7	1.2
FBVR	0.1	0.8	1.2	0.8	0.8	1.2

1000 Monte Carlo replications. LB denotes 95% lower bound and UB denotes 95% upper bound. $\mathcal{L}_{it} \sim i.i.d.\ N(0,1)$. We set $\mu = -1.662$ to obtain 50% of unit treated.

Appendix D. Test of identifying assumption

Table D.1. Size of the Test Statistics, Number of rejections out of 100 cases

			k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
t = 1	p = 1	Mean	0.090	0.121	-0.084	0.014	0.129	-0.130	-0.035	-0.133	-0.035	0.017
		Variance	0.826	1.136	1.159	1.018	1.195	0.795	0.830	1.347	0.996	0.970
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	0.033	0.037	0.045	0.299	-0.038	-0.037	0.074	-0.171	0.069	-0.025
		Variance	1.381	1.142	1.115	0.944	1.144	1.125	1.272	0.991	1.071	1.217
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	0.049	0.131	0.083	0.089	0.086	-0.088	0.044	0.121	-0.083	-0.178
		Variance	1.069	1.125	0.941	0.731	1.135	1.397	0.806	1.082	1.022	1.432
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	-0.114	0.037	0.106	-0.128	0.152	0.057	-0.207	-0.119	-0.038	-0.047
		Variance	1.016	1.083	1.110	0.953	1.024	0.762	0.965	1.066	0.979	1.042
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	-0.006	0.003	0.082	0.113	-0.089	0.023	-0.066	0.196	-0.166	0.004
		Variance	0.803	0.979	0.809	1.052	1.107	0.971	1.160	0.830	1.075	1.324
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.150	0.045	-0.028	-0.014	-0.154	-0.078	0.050	0.090	0.032	0.016
		Variance	1.249	1.136	0.889	0.980	1.121	0.803	1.154	1.161	1.263	1.189
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	0.141	-0.031	0.058	-0.058	-0.016	-0.002	-0.068	-0.046	-0.045	0.091
		Variance	0.779	0.899	1.063	1.244	0.897	1.089	0.882	1.118	1.210	1.063
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.102	0.022	-0.071	0.149	0.103	-0.043	0.130	-0.078	0.081	0.097
		Variance	0.979	0.921	0.994	0.974	1.049	0.963	0.979	1.037	1.170	1.116
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	0.174	0.047	0.125	-0.205	0.000	-0.011	0.085	0.027	0.025	-0.007
		Variance	1.064	1.002	1.012	1.119	0.996	1.216	1.017	1.113	0.918	1.198
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	-0.057	0.013	-0.037	-0.018	-0.122	-0.077	0.022	0.072	0.132	0.117
		Variance	0.733	0.905	1.093	0.992	1.193	1.028	1.279	0.942	0.786	1.002
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.073	0.006	-0.063	-0.243	0.162	-0.059	0.023	0.016	-0.099	0.336
		Variance	0.968	1.223	1.052	1.024	0.966	1.008	1.142	1.059	1.042	1.125
		10%	0	0	0	0	0	0	0	0	0	0
t = 2	p = 1	Mean	0.057	-0.130	-0.106	0.149	0.078	0.069	0.081	0.013	0.042	-0.120
		Variance	1.147	1.174	1.058	1.017	0.852	0.966	1.064	1.032	0.783	0.968
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	3.023	-0.115	-0.111	-0.208	-0.201	-0.037	-0.212	-0.141	-0.133	-0.159
		Variance	1.278	0.821	0.989	1.219	0.989	0.987	0.819	1.293	1.147	1.142
		10%	0	0	0	0	0	0	0	0	0	0

	p = 3	Mean	-0.088	-0.009	-0.050	-0.024	0.033	-0.189	-0.048	0.126	-0.041	0.085
		Variance	0.811	0.945	1.418	1.272	1.329	1.078	0.946	1.136	1.129	0.900
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	-0.047	-0.131	-0.237	-0.160	0.077	-0.013	0.028	0.014	-0.041	-0.015
		Variance	1.043	0.950	1.101	1.316	1.453	0.808	0.923	0.948	0.980	1.020
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	0.136	0.025	-0.073	-0.144	0.101	0.021	-0.016	0.073	-0.070	-0.019
		Variance	1.178	1.078	1.164	1.047	1.006	0.921	0.917	1.024	1.385	1.238
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.066	0.034	0.076	0.152	-0.361	0.075	0.036	-0.149	-0.053	0.076
		Variance	0.938	0.926	1.009	1.131	1.173	1.071	1.147	0.985	1.179	1.280
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	0.042	0.022	0.083	0.013	0.138	-0.223	-0.059	0.096	0.156	0.087
		Variance	0.914	0.976	1.147	1.118	1.078	1.023	1.177	0.801	0.956	0.839
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	-0.041	-0.003	-0.014	0.047	0.119	-0.033	-0.199	-0.058	-0.090	0.123
		Variance	1.162	1.045	0.939	1.189	1.014	0.942	0.891	1.198	1.022	1.090
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	-0.015	0.002	0.004	-0.205	0.065	0.105	0.068	-0.042	-0.119	0.284
		Variance	1.188	1.134	0.964	1.158	0.957	0.973	0.924	1.063	1.085	1.148
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	0.130	-0.226	-0.153	-0.135	0.022	-0.060	0.142	0.154	0.020	-0.016
		Variance	1.055	0.887	1.229	1.294	0.754	1.232	1.080	1.066	1.218	1.013
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.041	0.262	0.030	0.145	-0.045	-0.172	-0.049	-0.004	-0.057	0.156
		Variance	0.845	1.076	0.990	1.079	1.258	1.332	1.061	1.224	1.321	1.245
		10%	0	0	0	0	0	0	0	0	0	0
t = 3	p = 1	Mean	-0.088	-0.125	-0.027	0.070	-0.008	-0.074	-0.007	0.028	0.021	-0.056
		Variance	0.914	1.192	0.926	1.154	1.068	1.122	0.967	1.383	1.104	0.811
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	0.128	-0.433	-0.079	0.064	-0.098	0.136	0.022	-0.090	-0.127	-0.052
		Variance	0.859	1.088	1.022	1.338	0.996	1.236	1.144	0.979	0.968	0.998
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	0.027	2.956	-0.264	-0.323	-0.198	0.024	-0.265	-0.096	-0.198	-0.238
		Variance	0.942	1.167	1.021	0.812	0.946	1.190	1.293	0.895	1.102	0.969
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	0.080	0.037	-0.048	-0.241	-0.012	0.053	-0.064	0.062	-0.169	0.056
		Variance	1.262	0.767	1.109	1.102	1.052	0.762	0.958	0.988	1.000	0.832
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	-0.096	-0.013	0.094	-0.166	-0.072	-0.042	-0.034	0.051	0.110	0.042
		Variance	0.897	1.237	1.277	0.982	1.061	0.905	1.157	1.051	0.997	1.088
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.018	-0.039	-0.061	0.016	-0.215	-0.104	-0.086	-0.286	0.052	-0.134

		Variance	0.882	1.161	1.160	0.862	0.954	0.994	0.851	1.042	0.737	0.955
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	-0.175	-0.130	0.247	-0.025	-0.035	-0.465	-0.080	0.130	0.135	-0.122
		Variance	1.162	1.249	0.925	1.270	0.942	0.936	0.967	1.000	0.971	1.247
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.045	0.016	0.041	-0.138	-0.155	-0.015	-0.190	-0.082	0.110	-0.135
		Variance	1.101	1.172	1.189	0.929	1.101	0.850	1.078	1.042	1.023	1.413
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	-0.130	0.025	-0.056	-0.192	0.064	0.040	0.105	-0.174	0.054	-0.009
		Variance	0.683	1.124	1.193	0.946	0.972	1.001	1.019	0.821	0.923	0.972
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	-0.029	0.030	-0.068	0.025	0.026	0.098	0.059	0.049	-0.145	0.104
		Variance	0.927	1.421	1.000	0.911	0.999	1.434	0.752	1.178	1.015	0.930
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.061	-0.083	0.045	0.151	-0.145	-0.061	-0.148	0.015	0.110	-0.091
		Variance	0.888	1.068	0.984	0.896	0.978	1.293	1.122	0.989	0.997	1.315
		10%	0	0	0	0	0	0	0	0	0	0
t = 4	p = 1	Mean	0.084	0.207	0.083	-0.063	-0.114	-0.006	0.190	0.018	0.061	-0.159
		Variance	1.014	1.127	1.169	0.997	1.109	1.080	0.981	1.097	1.042	0.840
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	-0.038	-0.160	-0.268	0.026	0.103	0.007	0.196	0.025	0.055	-0.055
		Variance	1.062	1.012	1.304	1.253	1.002	1.281	1.335	1.087	0.815	1.031
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	-0.004	0.020	-0.067	0.143	0.101	-0.011	0.009	-0.048	-0.104	-0.114
		Variance	0.896	0.907	0.943	0.797	1.143	0.986	0.729	1.068	1.170	0.938
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	0.090	0.199	2.814	-0.137	-0.178	-0.087	-0.101	-0.274	-0.099	-0.119
		Variance	0.974	0.865	0.725	1.036	1.187	1.051	0.839	0.949	0.989	1.084
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	0.084	-0.190	0.027	-0.231	0.068	-0.159	0.039	0.098	-0.052	0.118
		Variance	1.035	0.986	1.513	0.722	1.247	1.033	1.328	1.116	1.190	1.041
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	0.144	-0.079	-0.203	-0.023	-0.034	-0.005	0.008	0.219	0.114	0.014
		Variance	1.017	1.096	1.236	1.208	0.926	1.236	0.889	0.997	1.068	1.375
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	-0.035	-0.064	0.137	-0.010	0.118	-0.205	-0.184	-0.038	0.187	-0.021
		Variance	0.905	0.831	0.778	1.093	0.990	1.017	1.148	1.123	1.045	1.053
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.063	0.049	-0.002	-0.001	-0.009	0.074	-0.161	0.099	0.123	-0.100
		Variance	1.051	1.152	1.022	0.997	1.136	0.992	1.231	1.022	1.081	0.887
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	0.174	0.098	-0.079	-0.048	-0.119	0.020	0.009	-0.071	-0.158	0.082
		Variance	0.937	0.897	1.000	1.086	1.165	0.814	0.939	1.177	1.171	1.290

		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	-0.018	-0.052	0.117	0.044	-0.036	-0.050	0.076	0.008	-0.172	0.196
		Variance	0.780	0.962	0.957	1.070	1.240	1.034	1.214	1.355	0.988	0.916
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.078	-0.087	-0.020	0.119	-0.084	-0.030	0.099	-0.016	0.111	-0.005
		Variance	0.968	1.164	0.980	0.923	1.096	0.872	1.146	1.029	0.772	1.356
		10%	0	0	0	0	0	0	0	0	0	0
t = 5	p = 1	Mean	0.121	0.146	0.018	-0.169	0.068	-0.077	-0.002	0.050	-0.061	-0.054
		Variance	1.084	1.120	1.118	0.807	0.906	1.228	1.036	1.243	1.104	1.036
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	-0.113	0.049	-0.038	-0.043	0.005	-0.126	0.032	0.085	-0.149	-0.080
		Variance	0.777	1.001	1.240	0.929	1.203	1.430	1.035	1.089	1.101	1.014
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	0.128	0.210	0.014	-0.344	-0.152	0.021	0.060	0.013	0.024	0.047
		Variance	0.889	1.110	1.139	0.910	1.219	1.065	1.202	0.970	0.958	1.144
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	-0.180	-0.004	0.122	-0.325	0.066	-0.059	0.081	-0.006	0.140	0.093
		Variance	1.226	0.905	1.152	1.257	1.040	1.197	1.075	0.974	0.886	0.981
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	0.040	-0.058	0.009	2.589	-0.304	-0.352	-0.186	-0.273	-0.019	-0.087
		Variance	1.209	1.219	1.515	0.791	0.793	0.836	0.801	1.012	1.209	1.128
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	0.042	0.042	0.049	0.141	-0.239	0.041	0.075	0.073	-0.057	0.126
		Variance	1.227	1.140	1.038	1.006	0.965	1.278	1.007	0.874	1.002	0.998
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	-0.211	0.080	0.155	0.072	0.013	0.002	0.085	0.109	0.056	0.158
		Variance	1.422	0.900	1.066	1.090	0.859	0.961	1.262	0.763	1.020	0.933
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.023	-0.056	-0.074	0.041	0.009	-0.010	-0.350	-0.178	-0.028	-0.063
		Variance	0.914	1.181	1.171	1.145	1.113	1.172	1.225	1.027	1.092	0.986
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	-0.011	0.113	-0.110	0.096	-0.014	0.007	0.033	-0.251	-0.125	-0.177
		Variance	1.050	0.986	1.327	1.157	0.930	1.051	1.176	1.280	1.030	0.971
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	0.081	0.040	-0.092	-0.037	-0.202	0.134	-0.098	-0.123	-0.062	-0.140
		Variance	1.605	0.986	1.173	1.180	1.166	1.194	0.930	1.068	0.927	1.182
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	-0.031	0.038	0.064	-0.171	-0.037	0.175	0.093	0.019	-0.020	-0.239
		Variance	1.028	1.295	0.920	0.987	1.014	0.827	0.949	0.962	0.902	1.304
		10%	0	0	0	0	0	0	0	0	0	0
t = 6	p = 1	Mean	-0.065	-0.061	0.131	0.098	-0.077	-0.009	0.030	0.069	0.022	-0.023
		Variance	0.990	0.899	1.062	1.178	0.931	0.910	0.931	1.131	1.118	0.828
		10%	0	0	0	0	0	0	0	0	0	0

	p = 2	Mean	0.000	-0.018	0.094	-0.010	-0.132	-0.015	-0.079	-0.053	0.087	0.118
		Variance	0.909	0.968	0.987	0.973	0.835	0.871	0.941	0.939	1.123	1.020
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	0.054	0.181	-0.145	0.083	-0.091	0.000	-0.105	0.169	-0.015	0.086
		Variance	0.859	1.062	1.249	0.997	1.362	1.113	1.134	1.052	0.801	1.056
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	0.052	0.177	-0.007	-0.028	-0.117	0.067	0.188	0.081	-0.020	0.143
		Variance	1.085	0.882	1.095	1.160	0.855	1.001	0.964	1.328	0.963	0.891
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	0.199	0.052	0.053	-0.045	-0.098	0.064	0.133	-0.009	-0.075	-0.131
		Variance	1.269	1.154	1.095	1.075	1.438	1.204	1.229	0.915	0.965	1.257
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.060	0.023	0.093	-0.021	2.660	-0.182	-0.178	-0.083	-0.151	-0.089
		Variance	0.966	1.250	1.024	0.739	0.700	0.984	1.031	1.059	1.235	0.915
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	0.033	-0.040	0.173	0.010	0.106	-0.126	0.077	0.050	0.164	-0.003
		Variance	0.931	0.918	1.267	1.078	1.083	1.171	0.800	1.044	1.150	1.211
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	-0.058	0.056	-0.112	0.179	-0.142	0.013	-0.125	0.008	-0.011	-0.041
		Variance	0.931	0.845	0.989	0.987	0.843	1.454	1.048	1.232	1.210	0.890
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	-0.056	-0.031	0.203	0.089	-0.085	0.146	0.135	-0.065	0.035	-0.100
		Variance	0.874	0.939	0.917	0.978	1.084	1.164	0.897	0.975	1.036	1.146
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	0.074	-0.186	-0.112	-0.004	-0.006	0.058	-0.085	-0.083	-0.030	-0.036
		Variance	1.053	0.660	1.159	0.925	1.123	1.267	1.006	0.867	0.832	0.939
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.015	-0.146	0.107	-0.029	-0.098	0.200	0.201	0.034	0.157	-0.225
		Variance	0.754	1.209	0.960	0.859	1.028	0.963	1.099	0.929	1.117	0.952
		10%	0	0	0	0	0	0	0	0	0	0
t = 7	p = 1	Mean	-0.102	-0.162	-0.089	0.093	-0.130	-0.061	-0.038	-0.037	0.223	-0.027
		Variance	1.284	0.948	1.205	1.208	1.024	1.013	0.944	1.037	1.246	1.070
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	0.005	0.099	-0.174	0.032	0.062	-0.173	0.133	-0.006	0.341	-0.135
		Variance	1.234	0.693	0.897	1.025	0.960	1.169	1.063	0.855	1.186	1.486
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	0.116	-0.017	0.041	0.063	0.022	-0.168	0.186	0.024	0.052	0.125
		Variance	0.904	1.303	0.834	1.140	1.170	1.168	1.043	1.122	0.833	0.988
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	0.121	-0.117	0.002	0.034	-0.198	-0.240	-0.061	0.148	0.007	-0.103
		Variance	0.936	1.057	0.949	1.082	1.224	1.108	0.922	1.020	0.906	1.287
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	-0.017	-0.062	-0.080	0.016	0.038	-0.150	-0.017	-0.051	0.122	0.041

		Variance	1.185	1.037	1.007	1.079	1.003	0.984	1.191	0.871	0.976	1.041
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	0.205	-0.061	-0.043	-0.022	0.124	-0.139	-0.019	-0.059	-0.093	-0.103
		Variance	0.563	0.923	1.117	0.946	1.151	1.140	1.118	1.038	1.044	0.937
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	0.236	0.135	0.052	-0.021	-0.080	2.715	-0.132	0.134	-0.172	-0.213
		Variance	0.962	1.217	0.928	0.921	0.907	0.913	1.063	1.267	0.955	1.129
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.066	0.045	-0.198	-0.018	-0.165	-0.062	-0.174	0.018	0.143	-0.062
		Variance	0.949	1.012	1.158	1.224	0.916	1.076	1.047	1.007	1.052	1.093
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	0.081	-0.013	0.027	-0.147	-0.084	-0.188	0.098	-0.354	0.158	-0.026
		Variance	1.148	1.220	0.782	1.267	0.727	1.160	1.057	0.991	1.178	1.015
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	-0.084	0.016	-0.023	0.040	0.088	0.000	-0.100	-0.099	-0.273	0.026
		Variance	1.094	0.890	1.058	0.815	0.787	0.986	1.096	1.137	0.902	0.986
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	-0.003	-0.082	0.105	-0.036	-0.167	0.075	-0.200	0.040	0.007	-0.229
		Variance	1.184	1.083	1.333	1.077	1.249	1.087	0.907	1.041	1.088	0.884
		10%	0	0	0	0	0	0	0	0	0	0
t = 8	p = 1	Mean	-0.154	-0.035	-0.136	-0.158	0.093	0.075	0.029	-0.118	-0.164	-0.064
		Variance	0.838	0.997	0.989	1.112	1.176	1.050	1.040	0.996	1.090	0.885
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	0.026	-0.066	-0.026	-0.113	0.021	-0.099	-0.210	-0.142	-0.012	0.046
		Variance	1.111	0.812	0.906	0.800	0.868	0.974	0.968	1.062	1.018	1.110
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	-0.039	-0.078	-0.086	-0.001	-0.026	0.067	-0.108	0.011	-0.029	-0.071
		Variance	1.401	1.087	0.964	0.918	1.055	1.097	1.180	0.895	1.116	0.906
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	-0.089	0.076	0.058	-0.099	0.133	0.227	0.085	-0.178	-0.042	-0.055
		Variance	0.888	1.034	0.922	1.318	1.089	1.469	1.115	0.833	1.123	1.047
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	0.290	-0.042	-0.077	0.077	-0.028	-0.124	-0.268	-0.149	0.016	-0.096
		Variance	1.114	1.120	1.033	1.035	1.100	1.359	1.019	0.980	1.075	0.991
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.077	0.156	-0.028	-0.045	0.105	0.033	-0.123	0.006	-0.099	-0.001
		Variance	1.100	0.856	0.955	0.945	0.844	1.096	1.144	0.877	1.312	1.180
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	-0.090	-0.056	-0.113	-0.027	0.177	0.268	-0.090	-0.082	0.145	-0.024
		Variance	1.037	1.060	1.086	1.152	1.061	1.082	1.071	1.291	1.150	0.910
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	-0.024	0.150	-0.177	-0.020	0.133	0.253	2.559	-0.128	0.222	-0.039
		Variance	1.015	0.931	0.994	0.885	1.007	1.115	1.189	0.857	0.810	0.818

		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	-0.069	-0.037	0.042	-0.004	0.016	-0.071	-0.067	-0.247	-0.057	-0.070
		Variance	0.958	1.008	0.914	0.924	1.196	0.996	0.934	0.983	1.226	1.144
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	0.109	-0.060	0.024	-0.114	-0.017	0.022	-0.153	0.064	-0.139	0.096
		Variance	0.871	0.986	0.707	1.081	0.977	1.065	0.989	0.820	0.931	1.023
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.069	-0.133	0.003	0.118	0.140	-0.099	-0.167	-0.003	-0.030	-0.177
		Variance	1.016	1.022	1.147	1.235	0.956	1.184	1.213	0.944	1.657	1.209
		10%	0	0	0	0	0	0	0	0	0	0
t = 9	p = 1	Mean	-0.118	-0.016	0.063	0.096	0.060	-0.004	-0.027	0.085	-0.025	0.243
		Variance	1.388	0.968	1.097	0.974	1.119	0.927	0.930	0.994	1.031	0.954
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	-0.166	-0.149	0.180	0.204	0.102	0.132	-0.151	-0.212	-0.018	0.083
		Variance	1.135	0.878	1.249	1.006	1.092	1.266	1.032	0.965	1.229	0.918
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	-0.052	0.074	0.075	-0.139	0.071	0.167	0.023	-0.133	0.172	0.087
		Variance	0.989	0.851	1.210	1.025	1.007	1.298	1.409	1.285	1.149	0.936
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	-0.192	0.118	0.105	-0.059	-0.148	0.021	-0.080	-0.147	0.176	0.041
		Variance	1.055	0.891	1.137	0.993	0.819	1.174	1.240	0.833	1.049	0.886
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	-0.071	0.003	-0.015	0.085	-0.008	0.073	0.124	-0.381	0.040	-0.098
		Variance	1.168	1.143	1.001	1.036	1.079	1.034	1.181	1.103	1.328	0.950
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.134	0.140	-0.029	-0.306	-0.068	-0.104	0.076	-0.256	0.019	0.048
		Variance	0.959	1.156	0.883	1.007	1.059	1.367	0.801	1.169	1.307	1.005
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	0.084	0.091	-0.008	0.048	0.055	0.143	0.011	-0.350	0.135	0.111
		Variance	1.276	1.005	0.901	0.887	1.040	0.987	1.160	1.211	0.951	0.964
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.071	-0.026	-0.064	-0.147	-0.082	0.051	-0.072	-0.075	-0.095	-0.064
		Variance	1.096	1.097	1.021	1.070	1.073	1.122	1.071	1.066	0.979	1.174
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	-0.064	0.026	0.113	0.049	-0.067	0.028	0.085	2.469	-0.122	-0.086
		Variance	1.035	0.966	1.126	1.316	1.004	1.163	1.112	1.240	1.010	0.907
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	0.021	-0.082	0.170	0.158	0.003	-0.168	-0.004	0.137	-0.184	0.275
		Variance	1.110	1.058	1.345	1.122	1.075	0.984	1.091	1.204	0.850	1.015
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.069	0.081	0.053	0.080	-0.011	0.066	-0.064	-0.020	-0.062	-0.071
		Variance	1.089	0.719	1.213	1.140	0.804	1.235	1.217	1.176	1.069	0.853
		10%	0	0	0	0	0	0	0	0	0	0

t = 10	p = 1	Mean	0.105	0.173	0.000	-0.032	-0.158	0.044	-0.059	-0.074	0.148	0.024
		Variance	1.395	1.108	0.998	1.186	1.061	0.988	1.173	0.783	0.873	1.296
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	-0.173	0.009	0.144	0.063	-0.046	0.038	0.102	-0.086	-0.071	0.110
		Variance	1.131	1.112	0.963	1.073	1.113	1.043	1.063	1.238	1.097	1.097
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	0.074	0.027	0.106	0.063	-0.090	-0.033	-0.002	-0.201	-0.008	-0.016
		Variance	1.013	0.813	0.946	0.919	1.134	1.217	1.080	0.977	1.233	1.287
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	-0.103	0.113	-0.026	-0.112	0.020	-0.133	0.146	-0.032	-0.149	0.028
		Variance	1.079	1.006	1.156	0.816	0.917	0.926	0.912	1.166	1.154	1.024
		10%	0	0	0	0	0	0	0	0	0	0
	p = 5	Mean	-0.044	-0.025	0.031	-0.056	-0.043	0.041	-0.015	-0.151	-0.237	-0.200
		Variance	1.011	1.122	1.304	1.052	1.676	1.208	1.073	0.780	1.138	1.021
		10%	0	0	0	0	0	0	0	0	0	0
	p = 6	Mean	-0.029	-0.017	0.068	-0.063	-0.034	0.019	0.013	0.021	-0.137	0.196
		Variance	0.883	0.955	1.139	0.857	1.043	0.880	0.881	1.057	1.032	1.387
		10%	0	0	0	0	0	0	0	0	0	0
	p = 7	Mean	-0.069	0.028	-0.039	-0.155	0.112	0.053	0.097	0.118	-0.151	0.174
		Variance	1.403	1.163	1.128	1.242	1.226	0.923	0.732	1.050	0.870	0.972
		10%	0	0	0	0	0	0	0	0	0	0
	p = 8	Mean	0.084	0.055	-0.157	0.116	-0.128	-0.021	0.069	-0.083	0.034	-0.131
		Variance	1.181	0.995	0.841	1.041	1.033	0.935	0.804	0.907	0.889	0.980
		10%	0	0	0	0	0	0	0	0	0	0
	p = 9	Mean	0.218	-0.112	-0.102	0.234	-0.100	-0.014	0.064	-0.089	0.046	0.121
		Variance	1.182	0.865	0.755	1.183	1.013	0.906	1.145	0.786	0.619	1.041
		10%	0	0	0	0	0	0	0	0	0	0
	p = 10	Mean	0.158	0.034	-0.098	-0.079	0.161	-0.004	0.080	0.102	2.478	-0.074
		Variance	1.058	1.160	1.011	1.326	1.197	1.188	1.073	1.155	0.925	0.796
		10%	0	0	0	0	0	0	0	0	0	0
	p = 11	Mean	0.007	0.029	0.100	-0.190	0.152	-0.233	0.090	-0.176	-0.043	-0.327
		Variance	0.994	0.809	1.093	0.848	0.975	1.014	0.901	1.106	1.133	1.124
		10%	0	0	0	0	0	0	0	0	0	0
t = 11	p = 1	Mean	0.069	-0.110	0.003	-0.087	0.047	-0.018	0.009	0.033	-0.020	0.035
		Variance	1.115	1.135	1.004	1.060	0.993	0.847	1.343	0.948	0.845	0.898
		10%	0	0	0	0	0	0	0	0	0	0
	p = 2	Mean	0.093	0.087	0.051	0.035	-0.088	-0.020	-0.148	0.076	-0.166	-0.344
		Variance	0.892	1.247	0.976	1.141	1.078	1.088	0.937	1.125	1.059	1.253
		10%	0	0	0	0	0	0	0	0	0	0
	p = 3	Mean	-0.033	-0.082	0.104	-0.026	0.084	-0.008	0.190	0.078	0.014	-0.133
		Variance	0.889	1.116	0.875	0.896	1.020	0.943	1.072	0.950	1.148	1.134
		10%	0	0	0	0	0	0	0	0	0	0
	p = 4	Mean	0.081	0.155	-0.104	0.024	-0.179	0.006	0.046	-0.051	0.146	-0.134

	Variance	0.978	1.188	0.962	1.056	0.994	1.082	1.235	1.098	1.074	1.160
	10%	0	0	0	0	0	0	0	0	0	0
p = 5	Mean	0.015	0.024	0.002	0.091	0.084	0.002	0.076	0.047	0.038	-0.204
	Variance	1.143	1.189	1.249	1.206	1.053	1.054	1.209	1.118	1.093	1.134
	10%	0	0	0	0	0	0	0	0	0	0
p = 6	Mean	0.113	0.040	0.008	0.045	0.087	-0.161	0.080	0.077	0.036	-0.121
	Variance	1.119	1.031	0.851	1.215	1.198	0.969	1.153	0.974	1.110	0.944
	10%	0	0	0	0	0	0	0	0	0	0
p = 7	Mean	-0.170	-0.079	0.059	0.135	-0.043	0.047	-0.014	0.161	-0.039	-0.125
	Variance	0.974	1.261	1.227	1.282	1.239	1.031	1.043	0.944	1.024	0.873
	10%	0	0	0	0	0	0	0	0	0	0
p = 8	Mean	-0.085	-0.127	-0.035	0.063	-0.165	-0.110	0.004	-0.088	-0.035	-0.089
	Variance	0.886	0.963	1.278	1.063	0.724	1.264	1.220	0.941	1.115	1.190
	10%	0	0	0	0	0	0	0	0	0	0
p = 9	Mean	0.017	0.175	-0.002	-0.139	0.109	0.107	-0.025	0.155	-0.124	-0.174
	Variance	1.029	1.034	1.044	0.915	0.926	1.007	1.117	0.881	0.959	0.881
	10%	0	0	0	0	0	0	0	0	0	0
p = 10	Mean	0.025	-0.093	-0.062	0.050	0.121	-0.081	0.068	-0.174	0.117	-0.052
	Variance	1.098	0.960	0.978	0.978	0.857	1.090	1.140	0.881	1.161	1.280
	10%	0	0	0	0	0	0	0	0	0	0
p = 11	Mean	0.203	-0.156	-0.019	-0.045	-0.032	0.008	0.086	0.023	-0.155	2.395
	Variance	1.041	1.128	0.879	0.928	1.114	0.982	0.920	1.112	0.994	1.165
	10%	0	0	0	0	0	0	0	0	0	0

Notes.
(i) $N = 1{,}000$, $\mathrm{T} = 11$, 100 replications.
(ii) ρ=0.4, γ=0.4.
(iii) We set $\mu = -1.662$ to obtain 50% of unit treated.
(iv) $\mathcal{L}_{\mathrm{it}} \sim \mathrm{i.i.d.\,N(0,1)}$.
(v) $\alpha = 1$.

Appendix E. Estimation with an alternative instrument

A natural test of robustness is to use an alternative instrument. In Table E.1, we replicate the estimation approaches presented in Table 6 using a dummy of a state Republican majority as an instrumental variable. We expect this dummy variable to influence the divesture decision since Republican politicians are reportedly more favorable to electricity restructuring (Joskow, 1997). This instrumental variable is also employed by Zhang (2007). The coefficient estimated by TSLS-FBVR, which equals -11, is lower but not statistically different from the coefficient obtained in our baseline estimation, which equals -7.6. However, standard TSLS and TSLS-probit provide very different point estimates in comparison to the results in Table 6. This further illustrates that TSLS and TSLS-probit are not reliable when the endogenous treatment is persistent. FVR appears to be more sensitive to the choice of instrumental variable than FBVR.

Table E.1. Estimation output of model (14) when an indicator for state Republican majority is used as instrument

Variable	TSLS	TSLS-probit	FVR	FBVR
$Divest_t$	-17.63**	-7.567	-19.125***	-10.959**
	(6.338)	(5.380)	(5.540)	(4.364)
Age_t	0.420	0.161	0.458*	0.248
	(0.277)	(0.229)	(0.263)	(0.228)
Age_t^2	-0.009**	-0.010**	-0.009**	-0.010**
	(0.004)	(0.004)	(0.004)	(0.004)
Year dummies	Yes	Yes	Yes	Yes
Treatment of obs. where UF=100	Dum. Var.	Dum. Var.	Dum. Var.	Dum. Var.
R^2	0.40	0.44	0.39	0.43
No. obs.	1851	1851	1851	1851

Notes: Dependent variable is *UF*. *UF* represents total number of outage hours divided by maximum potential generation hours. * p < 0.10, ** p < 0.05, *** p < 0.01. SE in brackets are robust to heteroskedasticity and autocorrelation with a Bartlett bandwidth = 2.

Appendix F. Spillover

Results in Section 5 rely on the assumption that untreated units are completely unaffected by the treatment of other units. In this section, we evaluate the reasonableness of this assumption. Spillover effects, i.e. untreated units are affected by treated units, can occur since information can flow across units directly as a result of joint stakeholders, and indirectly through industry associations and labor movements. [8] If spillovers are present, we shall under-estimate the effect of the treatment. In this section, we base our analysis on the FBVR estimator presented in Table 6 using the share of industrial electricity consumption as an instrument.

We present the results in Table F.1. The existence of spillover effects is tested at three different levels: (*i*) for nuclear reactors operated by the same operators (column 1), (*ii*) for nuclear reactors with similar technological characteristics (column 2)[9] and (*iii*) for nuclear reactors located in the same state (column 3) or in the same and neighboring states (column 4). We add the appropriate dummy variables to our base specification for each scenario.

Since none of the coefficients of these dummy variables is found to be significantly different from zero, we cannot reject the null hypothesis that there are no spillover effects from divested to non-divested reactors. This suggests that divestiture may lead to operational or managerial changes that are difficult to transfer to non-divested reactors. It is worth noting that this result contrasts with the findings of Craig and Savage (2013) who identify significant spillovers for thermal power plants in the U.S. following restructuring. This may be explained by nuclear reactors' complexity and specific regulations that make it difficult to transfer experience across reactors.

[8] For example, the Institute for Nuclear Power Operation fosters exchange of knowledge and experience across nuclear operators.

[9] We define reactor technology classes based on reactor containment type, steam system supplier and design type using data for the US Nuclear Regulatory Commission Information Digest 2012–2013. Available at: www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1350/appa.xls

Table F.1. Estimation output of model (14) and spillovers

Variable	Operational Spillovers (1) Mean (SE)	Technical Spillovers (2) Mean (SE)	Geographic Spillovers (3) Mean (SE)	(4) Mean (SE)
$Divest_t$	-7.716*** (2.266)	-7.189 ** (3.280)	-8.368 ** (3.958)	-9.447 (6.596)
$Op.Spill.Divest_t$	-1.275 (1.596)			
$Tech.Spill.Divest_t$		-0.720 (2.126)		
$Geo.Spill.Divest_t$			-0.862 (2.313)	-1.834 (4.700)
Age_t	0.169 (0.210)	0.144 (0.344)	0.200 (0.253)	0.259 (0.361)
Age_t^2	-0.009** (0.003)	-0.010*** (0.004)	-0.009** (0.004)	-0.009** (0.004)
Year FE	Yes	Yes	Yes	Yes
Treatment of obs. where UF=100	Dum. Var.	Dum. Var.	Dum. Var.	Dum. Var.
R^2	0.435	0.435	0.397	0.434
No. obs.	1851	1851	1851	1851

Notes: Column (3) limits geographical spillovers to reactors within the same state, and Column (4) allows divested reactors to influence reactors both within the same state and in neighboring states. * p < 0.10, ** p < 0.05, *** p < 0.01. UF represents total number of outage hours divided by potential generation and is the dependent variable. SE are robust to heteroskedasticity and autocorrelation with a Bartlett bandwidth = 2.