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

Michel Berthélemy, Petyo Bonev, Damien Dussaux and Magnus Söderberg

February 2017

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its second phase started in 2013 and there are five integrated research themes:

1. Understanding green growth and climate-compatible development
2. Advancing climate finance and investment
3. Evaluating the performance of climate policies
4. Managing climate risks and uncertainties and strengthening climate services
5. Enabling rapid transitions in mitigation and adaptation

More information about the Centre for Climate Change Economics and Policy can be found at: www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has nine research programmes:

1. Adaptation and development
2. Carbon trading and finance
3. Ecosystems, resources and the natural environment
4. Energy, technology and trade
5. Future generations and social justice
6. Growth and the economy
7. International environmental negotiations
8. Modelling and decision making
9. Private sector adaptation, risk and insurance

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: www.lse.ac.uk/grantham.

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

Michel Berthélemy[1], Petyo Bonev[2], Damien Dussaux[3], and
Magnus Söderberg[4]

[1]*Commissariat l'Energie Atomique et aux Energies Alternatives (CEA) - Université Paris-Saclay*
[2]*Mines ParisTech, PSL - Research University, CERNA - Centre for Industrial Economics,
i3, CNRS UMR 9217*
[3]*Grantham Research Institute on Climate Change and the Environment, London School of Economics.*
[4]*University of Gothenburg*

February 2017

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 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 Classification: 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 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

---

[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 favour of the status quo (Fernandez and Rodrik, 1991).

[2] Examples and references are provided in section 3 .

been used earlier by Honoré and Tamer (2006) and Carro (2007a). The FVR and FBVR approaches perform substantially better than both OLS and TSLS: the 95% confidence intervals of FVR/FBVR are up to 6 times smaller than the 95% confidence intervals of the TSLS and the empirical bias is up to 6 times smaller than the empirical OLS bias. The efficiency gap between FVR/FBVR and TSLS increases linearly with the number of periods. These results are consistent with our analysis of the problem of persistent treatment and are robust to the 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 a local average treatment effect estimator, 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%.

2 Related literature

The literature on weak instruments in a TSLS context is still growing, see Stock et al. (2002) for an overview of some of these papers. We contribute to this literature in two distinct ways. First, we formally describe a new setting that leads to a weak instrument. A distinctive characteristic of the persistent treatment problem is that the instrument becomes weak(er) over time. Intuitively, after the point in time of treatment of an individual, the treatment variable and the instrument become independent. The higher the number of time periods, the more severe the problem. Second, we suggest an enhancement of the instrument in such a setting through a modification of the first stage. The two closest related papers are those of Ratkovic and Shiraito (2014) and Kuersteiner and Okui (2010). Both of these papers suggest methods for improving the performance of the instrument(s). Their settings, however, differ substantially from ours. Ratkovic and Shiraito (2014) consider a case in which the instrument is weak because some agents are not influenced by the instrument (non-compliers). The authors tackle the problem by modelling non-compliance

and down-weighting those observations. In our case, however, the weakness of the instrument evolves over time due to the persistence of the treatment and not due to non-compliance. Kuersteiner and Okui (2010) consider a setting with many instruments and suggest to average over the first stage predictions produced by different sets of instruments. In our setting, however, the model is just-identified so this approach is not applicable.

3   The problem of persistent treatment

Suppose we have panel observations on a binary random variable D_{it} (treatment) and on an outcome variable Y_{it}. The index i indicates the cross sectional unit with $i = 1, \ldots, n$, and the index t indicates a 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-level productivity or individual wage. In many cases, the treatment variable D_{it} is endogenous due to unobserved selection into (or out of) treatment. Suppose we observe a period-specific instrument Z_{it} for D_{it}. Using the exogenous variation of the instrument, it is often possible to identify the causal effect of D_{it} on Y_{it}, 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, there is a persistent treatment problem if the following three conditions are fulfilled:

- The treatment is endogenous,

- The treatment is persistent: once a unit is treated, the treatment variable does not change its value for many, or all, subsequent periods,

- The instrument 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 (1992), Bertrand (2004), and exposure to oil shocks Raphael and Winter-Ebmer (2001). As a result of the three features listed above, the higher the number of periods of observation T is, the weaker the instrument.

In the context of GMM (and related approaches), a weak instrument might lead to a variety of estimation problems that are well known in the literature, Stock et al. (2002). In the just-identified case, the asymptotic variance is very high. As a result, the confidence intervals will be uninformative. Moreover, the confidence intervals may not have the correct nominal coverage, Staiger and

Stock (1997). In the over-identification case, the use of many weak instruments might in addition lead to a large bias of the estimator.

In this paper, we consider only the just-identified case. The main reason is that our approaches for strengthening the instrument impose a natural upper bound of the number of instruments that can be used (see section 6.1 for details).[3] In the just-identified case, GMM is known to be approximately unbiased in small samples. The main concern in this setting is the large variance of the GMM estimator, which results from the weak instrument.

We first formalise the problem of persistent treatment in a simplified framework. Generalisations to more complex settings are discussed at the end of the section.

Consider the panel model

$$Y_{it} = \alpha D_{it} + U_{it}, \tag{3.1}$$

where Y_{it}, D_{it} are defined as above and U_{it} is the unobserved disturbance of the model with $cov(D_{it}, U_{it}) \neq 0$. When Y_{it} represents production efficiency, U_{it} might capture maintenance routines that affect also market deregulation. Similarly, U_{it} might capture unobserved motivation that is correlated with the intention to obtain a master degree.[4] α is the causal effect of D_{it} on Y_{it} to be estimated. For ease of exposure, we assume that there are no observed covariates other than D_{it} and that Z_{it} is one-dimensional. Note that this paper is on fixed T panels with asymptotics over $N \rightarrow \infty$ (short panel). For the formalisation of the problem of persistent treatment, however, we will allow T to grow to infinity.

Definition: we call a binary treatment variable D_{it} a persistent treatment (PT) when

$$\lim_{t \to \infty} P\{D_{it} = 1\} = 1. \tag{PT1}$$

This definition is fulfilled for example in a setting in which i) each unit is treated eventually and ii) once treated, its treatment status remains unchanged. Although all our simulations and empirical examples assume both i) and ii), ii) is not necessary for the PT1 problem to exist and we do not assume it in our theoretical part in the next sections. Both i) and ii) are typical in policy reform settings, for which the number of periods of observation T is substantial (but still smaller than the number of cross sectional units n). If condition ii) is satisfied, then (PT1) is equivalent to the condition

$$\lim_{t \to \infty} P\{T_i \leq t\} = 1, \tag{PT2}$$

---

[3]In addition, many instruments might not be available in applications (which, in fact, is typically the case).
[4]Most panel models impose a one-way error model, $U_{it} = \mu_i + V_{it}$, where μ_i is an individual specific effect and V_{it} the idiosyncratic part of the disturbance, see for example Baltagi (2008).

where T_i is defined as the period in which individual i receives $D_{it} = 1$ for the first time. The random variable T_i is discrete and can be interpreted as the duration to treatment of individual i.

Empirical example: dynamic discrete choice To illustrate the relevance of our theoretical results, we use a dynamic discrete choice example throughout the paper. Dynamic discrete choice models have gained importance in econometrics in recent years, see e.g. Heckman and Navarro (2007) and Taber (2000). The decision to receive the treatment or not is modelled each period with an underlying utility threshold model. Suppose that for $t = 2, \ldots, T$, the treatment variable is generated according to the rule

$$D_{it} = \mathbb{1}\{-\theta Z_{it} + U_{it} < 0\} \quad \text{if} \quad D_{it-1} = 0 \quad \text{and} \quad D_{it} = 1 \quad \text{if} \quad D_{it-1} = 1. \quad (3.2)$$

D_{i1} is generated simply as $\mathbb{1}\{-\theta Z_{i1} + U_{i1} < 0\}$. The rule in the indicator function might depict the decision of a policy maker in period t that depends on some social welfare benefit θZ_{it} and some social cost U_{it}. A regulation is introduced in period t if the benefits are higher than the cost.[5] It is the participation of the structural regression error U_{it} in the first stage that makes the treatment variable D_{it} endogenous.

Define $Y_i = (Y_{i1}, \ldots, Y_{iT})'$, $Z_i = (Z_{i1}, \ldots, Z_{iT})'$, $D_i = (D_{i1}, \ldots, D_{iT})'$ and $U_i = (U_{i1}, \ldots, U_{iT})'$. We assume that **E1)** $Z_1, \ldots Z_n$ are independent, **E2)** $U_1, \ldots U_n$ are independent, **E3)** $Z_1, \ldots Z_n$ are jointly independent from $U_1, \ldots U_n$, **E4)** for each i and t, $Z_{it} \sim N(0, \theta^{-2}/2)$ and $U_{it} \sim N(0, 0.5)$, **E5)** and finally that for each i, $Z_{i1}, \ldots, Z_{iT}$ are independent and $U_{i1}, \ldots, U_{iT}$ are independent. Under the assumptions above, the term $-\theta Z_{i1} + U_{i1}$ follows a standard normal distribution. In this very simple setting, we can formulate the following lemma:

Lemma 3.1. *Under assumptions E1-E5, conditions* (PT1) *and* (PT2) *hold.*

The proof can be found in Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Condition (PT1) weakens the relation between the treatment variable and the instrument over time. Intuitively, after the point in time of treatment, the instrument and the treatment variable are independent. This intuition is captured by the following result:

Proposition 3.0.1. *Suppose that condition* (PT1) *is fulfilled. Furthermore, assume that the sequence* $(Z_{it})_{t=1,\ldots}$ *is stochastically bounded, i.e.* $Z_{it} = O_P(1)$. *Then*

$$\lim_{t \to \infty} cov(D_{it}, Z_{it}) = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.3)$$

---

[5]In the empirical evaluation section, we discuss a concrete example.

Empirical example: dynamic discrete choice (continued) Consider the dynamic discrete choice setting with D_{it} generated under the rule 3.2. Figure 1 depicts how the correlation ρ_{it} between D_{it} and Z_{it} decreases over time. Although this correlation is roughly 25% on a cross-sectional level, it decreases when further periods are added. For values of $T \geq 20$, this correlation remains within the interval $(-0.05, 0.05)$ and its empirical distribution is centered at 0.

[Insert Figure 1 about here]

Note that condition (3.3) is satisfied for any instrument whose variation does not "explode" over time, in particular also for a constant. The condition, however, will lead to a weak instrument problem only if the variation of the instrument does not decrease quickly with increasing t. In the simple setup of homoskedastic errors, this can be expressed with the following condition (we refer to it as **Weak Instrument due to Persistent Treatment (WIPT)**:

$$\sum_{t=1}^{T} \mathbb{E}[D_{it}Z_{it}] = o\Big(\sum_{t=1}^{T} \mathbb{E}[Z_{it}^2]\Big), \tag{3.4}$$

where the convergence is over $T \to \infty$.

To demonstrate the use of the WIPT condition, we assume that model (3.1) holds. The following Lemma holds:

Lemma 3.2. *Suppose that $(Y_i, D_i, Z_1)_{i=1,2\ldots}$ are i. i. d. with finite second moments and that $\mathbb{E}[Z_{it}U_{it}] = 0$ for all i and t. In addition, assume for the errors that $\mathbb{E}[U_{it}^2 \mid Z_{it}] = \sigma_u^2$. Then under the standard rank condition[6] $\mathbb{E}[D_i'Z_i] \neq 0$, it holds for the asymptotic variance $Avar(\hat{\alpha})$ of the (pooled) panel TSLS estimator $\hat{\alpha}$ under WIPT*

$$Avar(\hat{\alpha}) \to \infty \tag{3.5}$$

as $T \to \infty$.

The proof is straightforward. Under the assumptions of the lemma, the asymptotic variance of the pooled estimator of α fulfills

$$Avar(\hat{\alpha}) = \sigma_u^2 \Big(\sum_{t=1}^{T} \mathbb{E}[D_{it}Z_{it}] \sum_{t=1}^{T} \mathbb{E}[Z_{it}Z_{it}]^{-1} \sum_{t=1}^{T} \mathbb{E}[D_{it}Z_{it}] \Big)^{-1} \tag{3.6}$$

$$= \sigma_u^2 \Big(\frac{\sum_{t=1}^{T} \mathbb{E}[D_{it}Z_{it}]}{\sqrt{\mathbb{E}[Z_{it}Z_{it}]}} \Big)^{-2} \to \infty.$$

For the first equality, see for example chapter 8 in Wooldridge (2002). The assumptions of the lemma, in particular the independence of the sample draws

---

[6]See assumption SIV.2 on page 186 in Wooldridge (2002).

over time, make it possible to use a pooled estimator. We adopt this assumption together with homoskedasticity because then the variance of the pooled estimator has a very simple representation. These assumptions are however not necessary for our theoretical part in the next sections. The WIPT condition can be interpreted in the following way. The covariance between the instrument and the treatment (the l.h.s. of (3.4)) decreases much faster than the variance of the instrument (r.h.s. of (3.4)) over time .

If $Z_{i1}, \ldots, Z_{iT}$ are identically distributed, or at least have the same second moment μ_Z^2, then the WIPT condition is equivalent to the condition

$$\sum_{t=1}^{T} \mathbb{E}[D_{it} Z_{it}] = o(\sqrt{T}). \tag{WIPT2}$$

Empirical example: dynamic discrete choice (continued) We now demonstrate, that the (WIPT2) condition is fulfilled in the simple dynamic discrete choice setting we constructed. The following proposition holds.

Proposition 3.0.2. *Under assumptions E1-E5, condition* (WIPT2) *is fulfilled.*

Corollary 3.0.1. *Under assumptions E1-E5, the asymptotic variance Avar($\hat{\alpha}$) of the pooled TSLS estimator grows to infinity as $T \to \infty$:*

$$\lim_{T \to \infty} Avar(\hat{\alpha}) = \infty \tag{3.7}$$

$\square$

Remarks. 1. Although we use the TSLS estimator throughout the paper, in the just-identified case (the case of concern in this paper) all GMM versions, and in particular the STSLS and the system IV estimator (see p. Wooldridge (2002), are equivalent. Furthermore, the GMM and Limited Information Maximum Likelihood (LIML) Estimator are numerically equivalent. **2.** It is straightforward to generalise the WIPT (and WIPT2) condition to a setting with observed covariates, heteroscedasticity and dependent (over time) sample draws. Moreover, the condition holds also with demeaned variables as in the fixed effects TSLS model. Thus, and taking the previous remark into account, the persistent treatment problem holds for a very broad class of empirical settings.

4 Two approaches for strengthening the instrument in the context of persistent treatment

We now present two approaches for strengthening the instrument in the context of a PT. Both approaches remove unnecessary variation of the instrument.

4.1 Model and Notation

Consider the following linear model,

$$Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}, \tag{4.1}$$

where X_{it} is a $1 \times (K-1)$ dimensional random vector of observed individual characteristics, C_i is unobserved and time-constant and U_{it} is the unobserved error term. We allow for two types of endogeneity. First, we allow for $corr(X_{it}, C_i) \neq 0$. Second, as in the previous section, $corr(D_{it}, U_{it})$ is allowed to be different from zero. 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 $\mathscr{Z}_{it} = (\mathscr{Z}_{it,1}, \mathscr{Z}_{it,2}, \ldots, \mathscr{Z}_{it,M})$ that is exogenous and can be used as an instrument for the endogenous treatment D_{it}. For now, we put no restrictions on M. The analysis in section 6.1, however, reveals that our FVR and FBVR approaches do not allow for M being greater than 2. Write $W_{it} := (D_{it}, X_{it})$ and $Z_{it} := (X_{it}, \mathscr{Z}_{it})$. Furthermore, define $\bar{Y}_i := \frac{1}{T}\sum_{t=1} Y_{it}$ and $\widetilde{Y}_{it} := Y_{it} - \bar{Y}_i$ (and with analogous notation for all other random variables). The demeaned model 4.1 is

$$\widetilde{Y}_{it} = \alpha \widetilde{D}_{it} + \widetilde{X}_{it}\beta + \widetilde{U}_{it}, \tag{4.2}$$

or, equivalently,

$$\widetilde{Y}_{it} = \widetilde{W}_{it}\gamma + \widetilde{U}_{it}, \tag{4.3}$$

where $\gamma = (\alpha, \beta')'$. Finally, using matrix notation, model 4.3 can be written as

$$\widetilde{Y}_i = \widetilde{W}_i\gamma + \widetilde{U}_i, \tag{4.4}$$

where $\widetilde{Y}_i = (\widetilde{Y}_{i1}, \ldots, \widetilde{Y}_{iT})'$ and analogously for $\widetilde{W}_i$ and $\widetilde{U}_i$. Throughout the paper we assume that the observations (Y_i, W_i, Z_i) are i.i.d..

The standard approach is to use a pooled TSLS method to estimate γ. The standard assumptions are:

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

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

3 : $rank(\mathbb{E}[\widetilde{Z}'_i\widetilde{W}_i]) = rank(\sum_{t=1}^{T} \mathbb{E}[\widetilde{Z}'_{it}\widetilde{W}_{it}]) = K$.

Assumption 1 states the strict exogeneity of the instrument. Assumption 2 states that none of the elements of Z_{it} is a linear combination of the other elements. Assumption 3 is a rank condition that states that the instrument and the endogenous regressor are sufficiently related. Under assumptions 1-3, the parameter γ of the structural equation (4.1) is identified.

9

4.2 The FVR approach

The FVR consists of two steps. First, the instrument is adjusted in the following way. For each cross sectional unit, the values of the instrument for all periods after the period of treatment (this is the first period t in which the treatment variable D_{it} for an individual i takes value 1) are set equal to the value of the instrument in the period of treatment. In a second step, TSLS FE is performed with the adjusted instrument. The simple example in Table 1 illustrates this procedure.

[Insert Table 1 about here]

The second column contains the dependent variable, for example the percentage of operating hours in a certain period of time. The third column contains the values of the treatment variable, for example 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 other subsequent periods (here only period 4) to create the transformed instrument.

For a formal definition, we introduce the following additional notation. Let T_i be the period at which agent i is treated for the first time, $i = 1, \ldots, n$, and let $T_i > 1$ for all i. Let $P_i = P(T_i)$ denote the random $T \times T$ transformation matrix which 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 a 1 as elements p_{kl} for which $k \geq l$ and a 0 elsewhere. The columns $T_i + 1 \ldots, T$ consists 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 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 be easily 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 with $P(T_i)$ transforms the vector $(a_1, \ldots, a_T)'$ into the vector $(a_1, \ldots a_{T_i-1}, a_{T_i}, a_{T_i}, \ldots, a_{T_i})'$. Further, define the (deterministic) time-demeaning 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 with 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 FVR estimation approach can now be formally described in the following way.

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

Step 2: Estimate the equation

$$Q_T Y_i = Q_T W_i \gamma + Q_T U_i \tag{4.5}$$

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

$$\widehat{\gamma}_{FVR} := \Big[\big(\sum_{i=1}^{n} \widetilde{W}_i' \widetilde{\ddot{Z}}_i\big)\big(\sum_{i=1}^{n} \widetilde{\ddot{Z}}_i' \widetilde{\ddot{Z}}_i\big)^{-1}\big(\sum_{i=1}^{n} \widetilde{\ddot{Z}}_i' \widetilde{W}_i\big) \Big]^{-1} \big(\sum_{i=1}^{n} \widetilde{W}_i' \widetilde{\ddot{Z}}_i\big)\big(\sum_{i=1}^{n} \widetilde{\ddot{Z}}_i' \widetilde{\ddot{Z}}_i\big)^{-1}\big(\sum_{i=1}^{n} \widetilde{\ddot{Z}}_i' \widetilde{Y}_i\big),$$
$$\tag{4.6}$$

where $\widetilde{Y}_i = Q_T Y_i$ (with analogous notation for all other variables).

4.3 The FBVR approach

Similar to FVR, the FBVR transforms the instrument and then performs standard TSLS FE estimation procedure with the transformed instrument. The instrument is transformed in the following way. For all treatment periods ($D_{it} = 1$), the values of the instrument in FBVR and FVR are equivalent. For all periods prior to the treatment, the values of the instrument in FBVR 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.

[Insert Table 2 about here]

Formally, let the random matrix Q_i has 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 $\mathring{\mathscr{Z}}_i := Q_i \mathscr{Z}_i$ for $T_i \leq T$ and $\mathring{\mathscr{Z}}_i := \mathscr{Z}_i$ for $T_i > T$, and and write $\mathring{Z}_i := (X_i, \mathring{\mathscr{Z}}_i)$. Then the FBVR estimator is defined as

$$\widehat{\gamma}_{FBVR} := \Big[\big(\sum_{i=1}^{n} \widetilde{W}_i' \widetilde{\mathring{Z}}_i\big)\big(\sum_{i=1}^{n} \widetilde{\mathring{Z}}_i' \widetilde{\mathring{Z}}_i\big)^{-1}\big(\sum_{i=1}^{n} \widetilde{\mathring{Z}}_i' \widetilde{W}_i\big) \Big]^{-1} \big(\sum_{i=1}^{n} \widetilde{W}_i' \widetilde{\mathring{Z}}_i\big)\big(\sum_{i=1}^{n} \widetilde{\mathring{Z}}_i' \widetilde{\mathring{Z}}_i\big)^{-1}\big(\sum_{i=1}^{n} \widetilde{\mathring{Z}}_i' \widetilde{Y}_i\big),$$

4.4 Identification

The cost of strengthening the instrument in both approaches is a potential endogeneity problem. Both $\ddot{Z}_i$ and $\mathring{Z}_i$ depend on the time to treatment T_i, and T_i depends in a complex, non-linear way on U_i. As a consequence, stronger assumptions are needed to ensure identification. We adopt the following assumptions:

R1 $\mathbb{E}[U_{it} \mid Z_{i1}, Z_{i2}, \ldots Z_{iT}, C_i, T_i = k] = 0$ for $t = 1, \ldots, T$ and $k = 1, \ldots, T$

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

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

R3 a) $rank(\mathbb{E}[\widetilde{\ddot{Z}}_i' \widetilde{W}_i]) = rank(\sum_{t=1}^{T} \mathbb{E}[\widetilde{\ddot{Z}}_{it}' \widetilde{W}_{it}]) = K.$

R3 b) $rank(\mathbb{E}[\widetilde{\mathring{Z}}_i' \widetilde{W}_i]) = rank(\sum_{t=1}^{T} \mathbb{E}[\widetilde{\mathring{Z}}_{it}' \widetilde{W}_{it}]) = K.$

R1 is the new stronger assumption. It resembles the standard orthogonality condition in the TSLS panel framework, which is implied by the standard strict exogeneity restriction. The main difference is that in R1 the expectation contains a function that depends on the treatment point in time. Contrary to the strict exogeneity assumption, however, assumption R1 has no intuitive interpretation that can be directly linked to economic theory. Below we deal with this disadvantage in two different ways. First, in section 5, we discuss low-level assumptions in a particular context that imply R1. These low-level assumptions are easier to relate to economic intuition. Second, we test R1 in a simulation study in the framework of dynamic discrete choice. The results of the simulation reveal that R1 is a valid assumption in this context. The discussion of low-level assumptions and the simulation results can be used in applied research as a basis of decision whether it is plausible to use our approaches. The main result of this section is the following proposition:

Proposition 4.1. *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.

4.5 Asymptotic properties

Define $A = \mathbb{E}[\widetilde{W}_i'\widetilde{Z}_i]$, $B = \mathbb{E}[\widetilde{Z}_i'\widetilde{Z}_i]^{-1}$, $\Omega = \mathbb{E}[\widetilde{Z}_i'\widetilde{U}_i\widetilde{U}_i'\widetilde{Z}_i]$, and
$G = (ABA')^{-1}AB\Omega BA'(ABA')^{-1}$ (Ω is assumed to be positive definite). We define
$\tilde{A}, \tilde{B}, \tilde{\Omega}$ and $\tilde{G}$ analogously but with $\widetilde{Z}_i$ replaced by $\mathring{\widetilde{Z}}_i$. In addition, we denote by
$(V)_{lm}$ the element of the matrix V with row index l and column index m.

We need the following additional assumptions.

R4 a) $\mathbb{E}[|\ (\widetilde{W}_i'\widetilde{Z}_i)_{km}\ |] < \infty$, $\mathbb{E}[|\ (\widetilde{Z}_i'\widetilde{Z}_i)_{lm}\ |] < \infty$ and $\mathbb{E}[|\ (\widetilde{Z}_i'\widetilde{U}_i)_{m1}\ |] < \infty$ for $k = 1, \dots, K, l = 1 \dots, M$ and $m = 1, \dots, M$.

R4 b) $\mathbb{E}[|\ (\widetilde{W}_i'\mathring{\widetilde{Z}}_i)_{km}\ |] < \infty$, $\mathbb{E}[|\ (\mathring{\widetilde{Z}}_i'\mathring{\widetilde{Z}}_i)_{lm}\ |] < \infty$ and $\mathbb{E}[|\ (\mathring{\widetilde{Z}}_i'\widetilde{U}_i)_{m1}\ |] < \infty$ for $k = 1, \dots, K, l = 1 \dots, M$ and $m = 1, \dots, M$..

R5 a) $\mathbb{E}[|\ (\widetilde{Z}_i'\widetilde{U}_i)_{m1}\ |^2] < \infty$ for $m = 1, \dots, M$.

R5 b) $\mathbb{E}[|\ (\mathring{\widetilde{Z}}_i'\widetilde{U}_i)_{m1}\ |^2] < \infty$ for $m = 1, \dots, M$.

Assumptions R4 and R5 are standard TSLS assumptions, see for example theorems 3.1 and 3.2 in Wooldridge (2002). Under the i.i.d. assumption, assumption R4 a) ensures that $n^{-1}(\sum_{i=1}^{n} \widetilde{W}_i'\widetilde{Z}_i)$, $n^{-1}(\sum_{i=1}^{n} \widetilde{Z}_i'\widetilde{Z}_i)$ and $n^{-1}(\sum_{i=1}^{n} \widetilde{Z}_i'\widetilde{U}_i)$ converge in probability to their sample counterparts (due to the Weak Law of Large Numbers). Similarly for assumption R4 b) with $\widetilde{Z}_i$ replaced by $\mathring{\widetilde{Z}}_i$. Assumption R5 a) ensures that $n^{-1/2}(\sum_{i=1}^{n} \widetilde{Z}_i'\widetilde{U}_i)$ converges in distribution to a normal distribution. We can now state the asymptotic properties of the FVR and the FBVR approaches.

Proposition 4.5.1. *Asymptotic properties*
Under assumptions R1-R4, it holds

$$\gamma_{FVR} \xrightarrow{p} \gamma \qquad and \qquad \gamma_{FBVR} \xrightarrow{p} \gamma. \tag{4.8}$$

If in addition R5 holds, then

$$\sqrt{n}(\gamma_{FVR} - \gamma) \xrightarrow{d} N(\mathbf{0}, G) \qquad and \qquad \sqrt{n}(\gamma_{FBVR} - \gamma) \xrightarrow{d} N(\mathbf{0}, \tilde{G}). \tag{4.9}$$

We omit the proofs of this proposition since it follows identical steps as in the proofs of propositions 5.1 and 4.2 in Wooldridge (2002) with Z_i replaced by $\widetilde{Z}_i$ ($\mathring{\widetilde{Z}}_i$). Inference can be performed via consistently estimating G and $\tilde{G}$. A consistent estimate of G can be obtained by replacing A, B and Ω by consistent estimates

(equivalently for $\tilde{G}$). The standard approach is to take the sample averages $\hat{A} = n^{-1} \sum_{i=1}^{n} \widetilde{W}'_i \widetilde{Z}_i$, $\hat{B} = \left(n^{-1} \sum_{i=1}^{n} \widetilde{Z}'_i \widetilde{Z}_i \right)^{-1}$ and $\hat{\Omega} = n^{-1} \sum_{i=1}^{n} \widetilde{Z}'_i \hat{U}_i \hat{U}'_i \widetilde{Z}_i$. Here, $\hat{U}_i$ is a residual obtained with a consistent estimator for γ. Under the assumptions of proposition (4.5.1) $\sqrt{n}(\gamma_{FVR} - \gamma) \xrightarrow{d} N(\mathbf{0}, \hat{G})$, where $\hat{G} = (\hat{A}\hat{B}\hat{A}')^{-1}\hat{A}\hat{B}\hat{\Omega}\hat{B}\hat{A}'(\hat{A}\hat{B}\hat{A}')^{-1}$. The resulting estimator is robust to heteroskedasticity and serial correlation across time.

Remark. In some cases, the i.i.d. assumption on the cross-sectional units might be violated. A typical example for cross-sectional dependence is when the observations belong to clusters such as schools or families. In such cases, adjustments are necessary to reflect the nature of the cross-sectional dependence. Typical methods to deal with clusters are explicit modelling of peer effects, Sacerdote (2011), or an adjustment of the standard errors to make them cluster-robust, see e.g. Cameron and Miller (2015) for a discussion. Both approaches can be applied directly in our setting in a straight-forward way.

5 Towards low-level assumptions that imply assumption R1

Assumption R1 is more difficult to evaluate than the standard orthogonality assumption ($\mathbb{E}[Z_{il}U_{ip}] = 0$ for all l and p). General low-level assumptions are hard to find due to the nonlinearity of the condition. In particular, $\mathbb{1}\{T_i = k\}$ depends in a nonlinear way on $(Z_{it}, U_{it})_{t \in \{1,\dots,T\}}$. We therefore focus on one particular model of the first stage, namely the dynamic discrete choice data generation process, introduced in the empirical example in section 3 and also used in the simulation study in section 7.

Suppose now that the first stage is generated according to the rule

$$D_{it} = \mathbb{1}\{\delta D_{it-1} + \theta Z_{it} + \rho U_{it} + \epsilon_{it} > 0\}, \tag{5.1}$$

where $t \geq 2$. For $t = 1$, we define $D_{i1} = \mathbb{1}\{\theta Z_{i1} + \rho U_{i1} + \epsilon_{i1} > 0\}$. The term U_{it} is as before the error from the second stage (the structural equation). The term ϵ_{it} is an unobserved exogenous variation. For simplicity, the instrument is one-dimensional and there are no other observed covariates. The persistence of the treatment is achieved by a very large parameter value δ. Setting $\delta = \infty$ gives the data generating process

$$D_{it} = \begin{cases} \mathbb{1}\{\theta Z_{it} + \rho U_{it} + \epsilon_{it} > 0\} & \text{if} & D_{t-1} = 0 \\ 1 & \text{if} & D_{t-1} = 1. \end{cases} \tag{5.2}$$

If in addition $\epsilon_{it} = 0$ and $\rho = 1$, then we obtain the specification of the constructed empirical example in section 3.

Additional assumptions. To make our task tractable, we adopt the following additional assumptions:

L1: $\mathbb{E}[Z_{it}] = 0$.

L2 a): For each $t \in \{1, \dots, T\}$, $Z_{it} \perp\!\!\!\perp (Z_{i1}, \dots Z_{it-1}, Z_{it+1}, \dots, Z_{iT})$

L2 b): For each $t \in \{1, \dots, T\}$, $U_{it} \perp\!\!\!\perp (U_{i1}, \dots U_{it-1}, U_{it+1}, \dots, U_{iT})$.

L3: The vectors $(Z_{it})_{t=1,\dots,T}$, $(U_{it})_{t=1,\dots,T}$, and $(\epsilon_{it})_{t=1,\dots,T}$ are jointly independent.

Discussion of the additional assumptions. L1 is adopted for simplicity of exposition. By assuming a zero-mean instrument in assumption L1, we essentially focus on the important feature of the instrument, namely its variation. In particular, an instrument Z_{it} and its zero-mean counterpart $Z_{it} - \mathbb{E}[Z_{it}]$ i) have the same variance, ii) lead to identical first-stage F-statistics and iii) lead to identical estimates from the structural equation. L1 can be achieved through a simple (mechanical) normalisation of the instrument on a cross-sectional level. The assumption is fulfilled for the demeaned instrument $\tilde{Z}_{it}$ in the fixed effects model when the instrument Z_{it} is either identically distributed over time or has at least the same mean in each period. Therefore, assumption L1 is not a strong assumption. It is fulfilled (per construction) in the empirical example in section 3.

Assumption L2 a) imposes that there is no serial dependence in the instrument. This assumption might be plausible in a context where the instrument is chosen to be an economic shock. If the instrumental value in a subsequent period is anticipated due to serial correlation, this could lead to a violation of the exogeneity assumption if the agents act upon the anticipation. As a consequence, preferred instruments are based on economic variables that are hard to predict with past values. One example is fluctuations in exchange rates due to speculations on the financial market, see Revenga (1992). Another often chosen method for constructing an instrument is to isolate the random component of variation by eliminating smooth changes caused by factors evolving over time. Cho et al. (2012) and Hoxby (2000) for example use random variations of population in an area. Such unpredictability is compatible with the notion of an "economic shock".

Assumption L2 b) imposes that there is no serial dependence in the regression error of the structural equation.[7] Its weaker version - no serial correlation - is the standard assumption in the Random Effects (RE) model and is imposed to show the efficiency of the RE estimator. In addition, no serial correlation

---

[7]Note that assumption L2 does not preclude serial dependence in the exogenous variation of the first stage $(\epsilon_{it})_{t=1,\dots,T}$.

is imposed in the FE model in combination with a GLS estimation procedure. Under assumption L2 b), all serial dependence in the total regression error is due to the unobserved fixed effect C_i. Although this assumption might be hard to justify in some settings, serial correlation is easily tested with conventional methods, see for example chapters 7.8.5 and 8 in Wooldridge (2002). If the null hypothesis of no serial correlation is not rejected, the researcher might be willing to conclude that there is no serial dependence as well. This conclusion would be particularly justified in the standard AR(1) model $U_{it} = \nu U_{it-1} + \xi_{it}$ with $\nu \in (-1, 1)$ and i.i.d. ξ_{it}. If the hypothesis is rejected, then there are still various ways to deal with serial correlation in order to satisfy L2 b). First, one might include additional lags of the exogenous independent variables $X_{it-1}, \ldots, X_{it-l}$, where l is some integer smaller than t. This is a typical method used in the literature to ensure dynamic completeness in the conditional mean, see e.g. chapter 7.8.3 Wooldridge (2002). Models that are dynamically complete do not have serial correlation. Intuitively, adding past observed information could potentially account for all the dependence of the dependent variable over time. Related approaches include to explicitly model the dependence over time, for example through time dummies (which represent aggregate influences in separate periods) or time trend. Papers of Wansbeek and Knaap (1999) and Alvarez and Arellano (2003) are examples of studies that impose time trend and serially uncorrelated disturbances. Finally, under additional assumption on the joint distribution of $(U_{i1}, \ldots, U_{iT})$ it is possible to rescale the model so that L2 b) is approximately true. In particular, assume that the variance of $(U_{i1}, \ldots, U_{iT})$ is given by the matrix Ω_U. Multiplying both sides of the structural equation by $\Omega_U^{-1/2}$ yields disturbances $(U_{i1}^*, \ldots, U_{iT}*)$ whose variance matrix is given by the identity matrix (this is the standard approach used to calculate the (infeasible) GLS estimator). Under the additional assumption of joint normal distribution, the serial uncorrelatedness yields serial independence.[8]

Assumption L3 is stronger than the usual strict exogeneity assumption. However, this differentiation is not important in applied work. To the best of the authors' knowledge, all empirical papers defend the validity of their exclusion restrictions by arguing that the instrument is independent from the regression error. Economic intuition is based on this independence and not on the zero-conditional-mean assumption.

We now use these assumptions to explore when R1 is fulfilled. First, note

---

[8]Since Ω_U is rarely known, one must estimate it in a first step.

that

$$\mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip}] = \mathbb{E}[\mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip} \mid \mathbb{1}\{T_i = k\}]]$$
$$= \mathbb{E}[\mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip} \mid T_i = k]P\{T_i = k\}$$
$$+ \mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip} \mid \mathbb{1}\{T_i = k\} = 0]P\{\mathbb{1}\{T_i = k\} = 0\}]$$
$$= \mathbb{E}[\mathbb{E}[Z_{il}U_{ip} \mid T_i = k]P\{T_i = k\} + 0P\{\mathbb{1}\{T_i = k\} = 0\}]$$
$$= \mathbb{E}[Z_{il}U_{ip} \mid T_i = k]P\{T_i = k\}.$$

Therefore, if $P\{T_i = k\} > 0$ for all k then assumption R1 is equivalent to the assumption

R1*: $\qquad \mathbb{E}[Z_{il}U_{ip} \mid T_i = k] = 0$ for all $l, p, k \in \{1, \dots, T\}$.

Under the persistent treatment assumption ($D_{it} = 1$ if $D_{it-1} = 1$) and the discrete choice first stage, we can write

$$\mathbb{E}[Z_{il}U_{ip} \mid T_i = k] = \mathbb{E}[Z_{il}U_{ip} \mid D_k = 1, D_{k-1} = 0]$$
$$= \mathbb{E}[Z_{il}U_{ip} \mid D_k = 1, D_{k-1} = 0, \dots, D_1 = 0]$$
$$= \mathbb{E}[Z_{il}U_{ip} \mid \delta D_{k-1} + \theta Z_{ik} + \rho U_{ik} + \epsilon_{ik} > 0,$$
$$\delta D_{k-2} + \theta Z_{ik-1} + \rho U_{ik-1} + \epsilon_{ik-1} \leq 0, \dots,$$
$$\delta D_1 + \theta Z_{i2} + \rho U_{i2} + \epsilon_{i2} \leq 0, D_1 = 0]$$
$$= \mathbb{E}[Z_{il}U_{ip} \mid \theta Z_{ik} + \rho U_{ik} + \epsilon_{ik} > 0,$$
$$\theta Z_{ik-1} + \rho U_{ik-1} + \epsilon_{ik-1} \leq 0, \dots,$$
$$\theta Z_{i2} + \rho U_{i2} + \epsilon_{i2} \leq 0, \theta Z_{i1} + \rho U_{i1} + \epsilon_{i1} \leq 0] =: A.$$

We can differentiate between the following cases.

- Case I: $l, p > k$. Due to assumptions L2 and L3, we obtain $A = \mathbb{E}[Z_{il}U_{ip}] = 0$.

- Case II: $l > k > p$ or $p > k > l$ or $l > k \geq p$ or $p > k \geq l$. If $l > k > p$, then

$$A = \mathbb{E}[Z_{il}U_{ip} \mid \theta Z_{ip} + \rho U_{ip} + \epsilon_{ip} \leq 0]$$
$$= \mathbb{E}[Z_{il}]\mathbb{E}[U_{ip} \mid \theta Z_{ip} + \rho U_{ip} + \epsilon_{ip} \leq 0] = 0,$$

  where the last equality is due to assumption L1. The cases $p > k > l$, $l > k \geq p$ and $p > k \geq l$ are treated analogously.

- Case III: $l, p < k$ (cases with less or equal are treated analogously). Then $A = \mathbb{E}[Z_{il}U_{ip} \mid \theta Z_{ip} + \rho U_{ip} + \epsilon_{ip} \leq 0, \theta Z_{il} + \rho U_{il} + \epsilon_{il} \leq 0]$.

Therefore, the task to find low-level assumptions that imply R1 reduce in this framework to finding conditions under which $A = 0$ in Case III. Although this is still a hard technical task (these conditions will depend on the joint distribution

of all 6 random variables), it has the advantage that the conditional expectation does not depend on the time to treatment T_i but solely on the instrument, the regression error and the unobserved exogenous variation. Our simulations in section 7 reveal that under symmetric distributions of these three components, R1* cannot be rejected. Symmetry of the distribution of the instrument can be verified empirically, and symmetry of the distribution of the error term can be at least discussed based on knowledge and/or assumptions about the error term. We leave the mathematical proof of this result for future research.

6 Discussion on alternative specifications and some limitations

6.1 Multiple instruments and multicollinearity

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

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

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

It holds for the transformed instruments

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

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

6.2 Distributed lag formulation

In this subsection, we investigate Distributed Lag (DL) formulations of the structural equation (4.1). Since the number of periods T is held fixed, the DL specification can only be a finite DL formulation. In a DL formulation, lagged values of the independent variable have a direct impact on the outcome

variable. We analyse two different specifications: i) a specification with lags of the treatment variable and ii) a specification with lags of X_{it}.

Specification *i)* can be written in the following form:

$$Y_{it} = \alpha_1 D_{it} + \alpha_2 D_{it-1} + \cdots + \alpha_k D_{it-k+1} + X_{it}\beta + C_i + U_{it}, \tag{6.1}$$

where k is a positive integer such that $t > k$. Including a lagged treatment variable allows for a "dynamic treatment effect" in the sense that the total effect of the treatment on the dependent variable can be split into different periods. A typical argument against a DL formulation is multicollinearity among the lagged explanatory variables. This problem is particularly relevant in the setting of a persistent treatment. To see this, assume that $T = 3$ and that one lag of the treatment variable is included in the structural equation. Further, assume that the distribution of T_i is uniform: $P\{T_i = k\} = 1/3$ for $k = 1, 2, 3$. Then D_{i2} is equal to one with probability $2/3$ and D_{i3} is almost surely equal to 1. Therefore, in two thirds of the cases, D_{i3} and D_{i2} will have the same value. Since both D_{i3} and D_{i2} are regressors of Y_{i3}, this will lead to a data-induced multicollinearity and a dramatic increase of the standard errors. If $P\{T_i < T\} = 1$, then D_{T-1} and D_T are both equal to 1 which leads to a perfect multicollinearity and the rank condition R3 is violated.

Specification *ii)* does not have to suffer from the same drawback. If the lags of the other observed covariates X_{it} are not highly correlated, then a DL specification can be estimated under assumptions R1-R5. Note that the orthogonality condition in R1 implies that past values of the observed covariates X_{it} are not correlated to current values of the error term, once the fixed effect C_i has been averaged out. A DL formulation therefore does not induce endogeneity of the lags.

6.3 Including time dummies

The question we analyse now is whether including time-specific effects into the main structural equation would lead to multicollinearity for similar reasons as described in 6.1. Let $\lambda_1, \ldots, \lambda_{T-1}$ be dummy variables with $\lambda_t = 1$ if the period of observation is t, and 0 otherwise, $t = 1 \ldots, T - 1$. Perfect multicollinearity arises when D_{it} is a linear combination of $\lambda_1, \ldots, \lambda_{T-1}$:

$$D_{it} = \delta_1 \lambda_1 + \cdots + \delta_{T-1} \lambda_{T-1} \tag{6.2}$$

Since, however, $\lambda_1, \ldots, \lambda_{T-1}$ are deterministic functions of the period of observation t, equation (6.2) implies that D_{it} is also a deterministic function of time. As a consequence, all units would have to be treated in the same period of observation t, a feature that we explicitly preclude in our model setting. This

conclusion is reached for the standard TSLS model as well, since it only considers the structural equations, not the first stage.

The first stage can be analyzed using equivalent arguments. The instrument in both our approaches is transformed as a function of the point in time of treatment T_i.[9] Therefore, perfect multicollinearity is only induced if there is no variation in T_i. Hence, incorporating time-specific effects into our setup does not violate the rank conditions (R2 and R3) and does not lead to perfect multicollinearity on the population level. We demonstrate this in the empirical investigation, where we include time-specific effects in the main specification.

Even if the rank condition on the population is not violated, incorporating time-specific effects might dramatically increase the standard errors if there is little variation in the point in time of treatment T_i *in the sample at hand*. Low variation in T_i in the sample will namely lead to a high *data-induced* multicollinearity of $(D_{it}, \lambda_1, \ldots, \lambda_{T-1})$.

6.4 A note on a first-differencing approach

In this section, we briefly discuss an alternative estimation approach often used in the context of panel data, namely that of first differencing (FD). For a random variable A_t, define ΔA_t to be the difference between A_t and its first time lag A_{t-1}: $\Delta A_t = A_t - A_{t-1}$. Then a first-differencing transformation on the main model (4.3) leads to the reduced form model

$$\Delta Y_{it} = \Delta W_{it}\gamma + \Delta U_{it}, \qquad t = 2, \ldots, T. \tag{6.3}$$

A standard FD approach in our context runs a pooled TSLS regression on the first-differenced data.

First, note that a FD approach alone cannot mitigate the persistent treatment problem. To see this, consider Table 3. It is built in the same way as Table 1 and uses the same numerical example. The difference is that the values in Table 3 are first-differenced. The table contains therefore only the periods 2, 3, and 4. The first-differenced treatment variable still has only two possible values 0 and 1 (-1 is not an admissible value because we do not allow the (persistent) treatment to change from 1 back to 0). Moreover, for all periods after T_i, the value of ΔD_{it} is equal to 0. The first-differenced instrument, on the other hand, is allowed to change values from period to period, because we have not placed any persistence assumptions on its increments (first differences) . Thus, the FD approach on its own must necessarily suffer from the WIPT problem.

[Insert Table 3 about here]

---

[9]Note, that in both FVR and FBVR approaches, only the exclusion restriction $\mathscr{Z}_{it}$ is transformed, not the included exogenous covariates X_{it}.

The FD transformation can be used, however, in a combination with the FVR and the FBVR approaches in the same way as the within transformation. The adjustment of the instrument has analogous steps, see for example the last column of Table 3 for the FVR-transformed instrument. Assumption R1 (in relation to assumptions R2 and R3) is sufficient for consistency (and identification) in the context of the FD transformation.[10] This can be seen in a straightforward way. In particular, consider the k-th element of the vector $\Delta \ddot{Z}_i := (\Delta \ddot{Z}_{i2}, \ldots, \Delta \ddot{Z}_{iT})$. It is equal to the expression (A.5) in the appendix (see proof of proposition 4.1), but with the term $-T^{-1}\sum_{l=1}^{T_i} Z_{il} - T^{-1}(T - T_i)Z_{iT_i}$ replaced by the term $-Z_{ik-1}\mathbb{1}\{T_i \geq k\} - Z_{iT_i-1}\mathbb{1}\{T_i < k\}$. All further steps of the proof of proposition 4.1 presented in the appendix are analogous. Together with assumptions R4 and R5, this implies the consistency and normality of the FVR and FBVR approaches with a FD transformed model. This result is not surprising, as it holds for the standard TSLS procedures as well, see p. 279 in Wooldridge (2002) for a discussion.

The FD is preferred to the standard within transformation when there are reasons to believe that the errors follow a random walk. In such cases, under a standard homoskedasticity-type assumption (see for example assumption FD.3 in chapter 10.3.1 of Wooldridge (2002)), the FD approach is more efficient than the within estimator. Because the manipulations of the FVD and the FBVR approach does not alter the form of the errors (only the instrument is transformed), this advantage of the FD approach carries over when FVR and FBVR are subsequently used.

6.5   A "local treatment" estimation approach

We analyse an approach, in which only observations just after and just before the point of treatment T_i are used.[11] Formally, suppose that i is held fixed. If $2 \leq T_i \leq T$, then the observations of the i-th unit from the periods $T_i - 1$ and T_i are kept while all other periods are excluded. If $T_i > T$ (or if the unit is never treated), or $T_i = 1$ (immediately treated), then the observations on unit i are not used. As an example consider Table 4.

[Insert Table 4 about here]

It represents a panel of length 4 with 4 observational units (O_1 to O_4). Each row represents one period of observation t. For each unit and each period of observation, the table contains a triple ($Y_{it}, D_{it}, X_{it}, Z_{it}$) of observed values for the

---

[10]With R2 and R3 appropriately modified: instead of the $\tilde{W}_i$ variable, one has to insert a ΔW_i variable, i.e. FD-transformed variable.

[11]We write "local treatment" in inverted commas to emphasize that this approach is not to be confused with the Local Average Treatment Effect (LATE) approach from the cross section literature.

dependent variable, the treatment variable, an exogenous observed covariate and the instrument (in that order). The units O_1 - O_3 are treated in the second, fourth, and the first period, respectively, and the fourth unit is not treated. In a first step, the "local treatment" approach would reduce this panel to the panel shown in Table 5. Only observations on units O_1 and O_2 around the treatment period are kept, i.e. periods 1 and 2 for unit O_1 and periods 3 and 4 for unit O_2. In a second step, a standard panel TSLS is performed.

[Insert Table 5 about here]

To compare this method with our proposed approaches (FVR and FBVR), we first note that assumption R1 is sufficient for the validity of the exclusion restriction in the "local treatment" approach. To see this, write

$$\mathbb{E}[Z_{iT_i}U_{iT_i}] = \sum_{t=1}^{T} \mathbb{E}[Z_{it_i}U_{it_i}\mathbb{1}\{T = t\}]. \tag{6.4}$$

When we pick up observations right before and right after the random treatment period T_i, the left hand side of (6.4) must be equal to zero in order to have a valid instrument. This is ensured when R1 holds, because R1 implies that the r.h.s of (6.4) is zero. Thus, R1 ensures identification and consistency.

One big disadvantage of the "local treatment" approach, however, is that the number of observations used for estimation is dramatically reduced. There are two reasons for this reduction: first, only units treated between the second and the last period of observation are included. The loss of observation increases with the share of untreated cross-sectional units. Second, only two observations are included for each included unit of observation. Thus, the number of lost observations is $T - 2$ per included unit. In the example above, the initial number of 16 observations is reduced to 4, leading to a 75% reduction of all observations. We demonstrate in the Monte Carlo section that such a loss of observations leads to a dramatic increase in MSE.

7 Monte Carlo Simulations

7.1 Data Generating Process

Our data generating process builds on previous work estimating dynamic binary choice models with unobserved heterogeneity (Honoré and Tamer, 2006; Carro, 2007b). It consists of a structural model (7.1) and a discrete choice model (7.2):

$$Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it} \tag{7.1}$$

$$D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\} \tag{7.2}$$

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 $U(0,1) \sim i.i.d.N(0,1)$.

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 $\mathscr{Z}_{it} \sim i.i.d.N(0,1)$ is used as an instrumental variable for the endogenous treatment D_{it}. The strength of the instrument $\mathscr{Z}_{it}$ is controlled by the size of the parameter θ. 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, $\epsilon_{it} \sim i.i.d.N(0,1)$ is a random exogenous 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 θ and ρ. To ensure that the level of endogeneity (instrumental strength) does not change when the value of $\theta(\rho)$ is modified, the parameter λ is set equal to $\sqrt{(1 - \theta^2 - \rho^2)}$. This allows us to keep $Var\left[\theta \mathscr{Z}_{it} + \rho U_{it} + \lambda V_{it}\right]$ equal to 1 when the value of θ or ρ changes. In our baseline simulation, we generate data for 1,000 units ($n = 1,000$) over 15 time periods ($T = 15$). We choose the value for μ so that approximately half of the units receive treatment during the observation period.

7.2 Simulation Results

We draw 1,000 samples for each of ten different empirical settings. The strength of the instrumental variable, captured by θ, is the only parameter that varies between the empirical settings. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\delta = 50$, $\mu = -1.663$, the exclusion restriction $\mathscr{Z}_{it}$ is normally distributed, and the share of units treated equal to 50%. Table 6 and Table 7 show the absolute empirical bias and the 95% confidence interval of the simulated α coefficient estimated by six different fixed-effect estimators: OLS, TSLS, Two Stage Local Treatment Estimator (TSLTE), TSLS-Probit, FVR, and FBVR. The Two Stage Local Treatment Estimator is a TSLS estimator using only the observations before and after the treatment time.[12] TSLS-Probit uses the predictions of a Probit model

---

[12]This is the estimation approach based on the strategy considered in section 6.5.

as an instrument for the treatment in a TSLS (Wooldridge Jeffrey, 2002). Table 6 also reports the Root Mean Square Error, the average first stage F-statistic, and the type II error frequency. The only difference between the two tables is the level of endogeneity equal to $\rho = 0.4$ in Table 6 and to $\rho = 0.6$ in Table 7.

Before looking at the relative performance of the five IV-estimators, we perform some consistency checks. Within Table 6 and Table 7, the absolute bias of the OLS estimate is the same in all empirical settings. This significant bias appears by design since the endogeneity parameter ρ is different from 0. The absolute bias of OLS is higher in Table 7 than in Table 6 since ρ is larger in Table 7. As expected, the distribution of $\hat{\alpha}^{OLS}$ does not change either as the strength of the instrument varies. As the level of endogeneity decreases and the strength of the instrument increases, the absolute bias of the five IV-estimators declines and their distributions tighten around the true value of α. The only exception is TSLTE whose bad performance is due to the large loss of observations as explained in Section 6.5. For clarity and because TSLTE is highly biased, we focus in the following on TSLS, TSLS-Probit, FVR, and FBVR. We will use TSLS to refer to both TSLS and TSLS-Probit estimators as they provide very similar results.

In almost all empirical settings, the absolute bias of FVR and FBVR is higher than the absolute bias of TSLS but has the same order of magnitude and is very small. However in every setting, FVR and FBVR are clearly more efficient than TSLS. More precisely, the 95% confidence interval of TSLS is 5 to 6 times larger than the 95% confidence interval of the FVR and FVBR. Importantly, this efficiency gap is the same for the different levels of strength of the original instrument θ.[13] Therefore, TSLS does not catch up FVR and FBVR in terms of efficiency as the original instrument gets stronger. Finally, FBVR performs slightly better than FVR in terms of bias while both estimators are equivalently efficient.

The 95% confidence bounds of TSLS are not only larger than the bounds of FVR and FBVR, they are also misleading. TSLS frequently fails to reject the null hypothesis while it is not true in our case. Their type II error frequency goes up to 80%.[14] Even if it diminishes with the strength of the instrument, the false inference remains frequent. Unlike TSLS, both FVR and FBVR almost never fails to reject the null hypothesis.[15] The higher efficiency of both FVR and FBVR is also reflected in the reported first stage F-statistics that are much larger than for

---

[13]Only for the lowest strength of the original instrument $\theta = 0.2$, the efficiency gap is between a factor 8 and 9.

[14]Note that this frequent error takes place even when the first stage F-statistics is higher than 10 which is the rule of thumb value used by applied econometricians.

[15]The highest type II error frequency for both FVR and FBVR equals 1.3 % and corresponds to the empirical setting having the weakest original instrument.

the competing IV estimators. These results are observed in both Table 6 and Table 7.

Finally, the FVR and FBVR performance relative to TSLS increases with the length of the panel. Table 8 and Table 9 both show the simulation results with different panel length T varying from 4 to 24.[16] The only difference between the tables is the endogeneity level ρ that is higher in Table 9 than in Table 8. In both tables, we obtain practically the same results. As T increases, the precision of TSLS decreases. This result corresponds precisely to the WIPT problem. In contrast, the precision of FVR and FBVR improves as T gets higher. The efficiency gap increases linearly with T and goes from 2 when $T = 4$ to 9 when $T = 24$. When the panel is short $T = 4$, FVR and FBVR still performs better than TSLS which erroneously fails to reject the null hypothesis 25% of the time. Note that the absolutes biases are not affected by variation in T.

In summary, we draw four 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 that frequently lead to type II errors, 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 the instrument $\mathscr{Z}$ is sufficiently related to the treatment D. Fourth, the efficiency gap between the FVR and FBVR approaches and TSLS increases linearly with the length of the panel. This is consistent with our analysis of the problem of persistent treatment.

We perform several robustness checks to verify the sensitivity of our main results. First, we allow the instrumental variable $\mathscr{Z}_{it}$ to be drawn from a non-normal distribution. Second, we generate samples where the share of units treated varies from 30% to 90%. Details of how these robustness tests were performed, as well as the results, are presented in Appendix B.

[Insert Table 6 to 9 about here]

7.3 Empirical test of the identifying assumption

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 the data generating process described by equation (7.1) and

---

[16]T and μ are the only parameters that vary within these tables. We change μ to keep the share of units treated close to 50% for each setting to make them comparable.

(7.2). Note that R1 implies the following assumption:

R1* $\qquad \mathbb{E}\left[\mathscr{Z}_{it} U_{ip} | T_i = k\right] = 0$ for all $t, k, p \in \{1, \ldots, T\}$.

The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathscr{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1*, the statistic $\mathscr{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp} / \sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$. Appendix C 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.

8 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.[17] 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.[18] 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.

---

[17]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.

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

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 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. Thus, these two studies illustrate two empirical traps: first that the instrument is weak, and second that the endogeneity problem is ignored.

8.1 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). Finally, data about state level political majority comes from the US census bureau.

Table 10 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 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.

[Insert Table 10 about here]

8.2 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 $UF = 100$, that is when reactors have not produced any electricity during a given year.

Moreover, we include year fixed effects η_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 or procedures that are influenced by technology, but that are unobserved in our data.

To estimate the effect of divestiture on the performance of nuclear reactos, we use the following model:

$$UF_{it} = \beta_0 Divest_{it} + \beta_1 Age_{it} + \beta_2 Age_{it}^2 + \beta_3 S_{it} + \eta_t + \lambda_i + \epsilon_{it} \tag{8.1}$$

where i is the reactor, t is the year and ϵ_{it} is the error term. 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 base-load 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).

28

The estimations of (8.1) 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.

[Insert Table 11 about here]

The main results are presented in Table 11. 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 standard error $SE(\beta_0)$ 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%.

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 11. The methodology and detailed results are given in Appendix D.

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 E. The results indicate that we cannot reject the null hypothesis that there are no spillover effects from the divestiture of other reactors.

9 Conclusion

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.

References

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics 27*(7), 1189–1204.

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*(4), 1121–1159.

Ando, A. W. and K. L. Palmer (1998). *Getting on the map: the political economy of state-level electricity restructuring*. Resources for the Future Washington, DC.

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

Baltagi, B. (2008). *Econometric Analysis of Panel Data*. John Wiley & Sons.

Bertrand, M. (2004). From the Invisible Handshake to the Invisible Hand? How Import Competition Changes the Employment Relationship. *Journal of Labor Economics 22*(4), 723–766.

Borenstein, S. and J. Bushnell (1999). An empirical analysis of the potential for market power in californias electricity industry. *The Journal of Industrial Economics 47*(3), 285–323.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

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

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

Cho, H., P. Glewwe, and M. Whitler (2012). Do reductions in class size raise students' test scores? Evidence from population variation in Minnesota's elementary schools. *Economics of Education Review 31*(3), 77 – 95.

Coate, S. and S. Morris (1999). Policy persistence. *American Economic Review 89*(5), 1327–1336.

Damsgaard, N. (2003). Why deregulation? economics and politics in retail electricity markets. *Deregulation and Regulation*, 1.

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

Delmas, M. and Y. Tokat (2005). Deregulation, governance structures, and efficiency: the us electric utility sector. *Strategic Management Journal 26*(5), 441–460.

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

Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual- specific uncertainty. *The American Economic Review 81*(5), 1146–1155.

Green, R. (1996). Increasing competition in the british electricity spot market. *The journal of Industrial economics*, 205–216.

Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Honoré, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica 68*(4), 839–874.

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

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

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics 115*(4), 1239–1285.

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

Kuersteiner, G. and R. Okui (2010). Constructing optimal instruments by first-stage prediction averaging. *Econometrica 78*(2), 697–718.

Raphael, S. and R. Winter-Ebmer (2001). Identifying the effect of unemployment on crime. *The Journal of Law & Economics 44*(1), 259–283.

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

Revenga, A. L. (1992). Exporting jobs?: The impact of import competition on employment and wages in u.s. manufacturing. *The Quarterly Journal of Economics 107*(1), 255–284.

Sacerdote, B. (2011). *Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?*, Volume 3 of *Handbook of the Economics of Education*, Chapter 4, pp. 249–277. Elsevier.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics 20*(4), 518–529.

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

Wansbeek, T. J. and T. Knaap (1999). Estimating a dynamic panel data model with heterogenous trends. *Annales d'Économie et de Statistique* (55/56), 331–349.

Warren, P. L. and T. S. Wilkening (2012). Regulatory fog: The role of information in regulatory persistence. *Journal of Economic Behavior & Organization 84*(3), 840–856.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge Jeffrey, M. (2002). Econometric analysis of cross section and panel data. *Cambridge, MA: Massachusetts Institute of Technology*, 623–625.

Zhang, F. (2007). Does electricity restructuring work? evidence from the us nuclear energy industry. *The Journal of Industrial Economics 55*(3), 397–418.
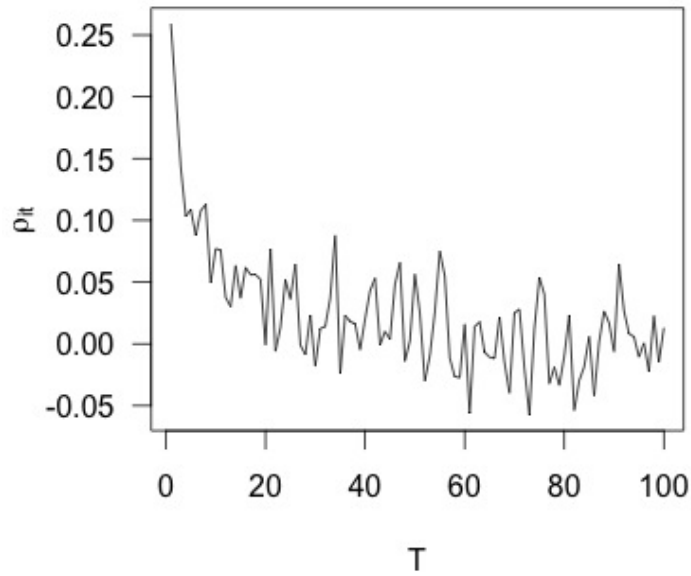
Figure 1: The correlation $\rho_{it} = corr(D_{it}, E_{it})$ on the y axis with (increasing) T on the x axis. The data generating process is the dynamic discrete choice described in the empirical example (assumptions E1- E5). θ is chosen to be 0.4. For each T, the sample size is 1000.

Table 1: FVR: a data example

Period	Y	D	Z	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

Table 2: FBVR: a data example

Period	Y	D	Z	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

Table 3: FD: a data example

Period	ΔY	ΔD	ΔZ	ΔZ_{FVR}
2	-1 %	0	-595	-595
3	6%	1	1787	1787
4	2%	0	-3486	0

Table 4: A "local treatment" approach: a data example

Period	O_1	O_2	O_3	O_4
1	(65,0,11, 140)	(59,0,13, 140)	(65,1,16, 210)	(71,0,14, 140)
2	(64,1,12, 210)	(65,0,13, 120)	(63,1,15, 150)	(69,0,11, 130)
3	(66,1,13, 170)	(62,0,11, 120)	(64,1,11, 160)	(67,0,13, 120)
4	(67,1,11, 140)	(67,1,12, 180)	(60,1,17, 165)	(68,0,18, 110)

Table 5: A "local treatment" approach: a data example continued. Transformed panel.

O_1	O_2
(65,0,11, 140)	(62,0,11, 120)
(64,1,12, 210)	(67,1,12, 180)

Table 6: main simulation results

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
1	$\theta = 0.2$	OLS	50	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.48	0.86	2.1	0.33	49.07	2.3
		TSLS	50	0.01	-2.26	4.23	2.74	10.91	82.9
		TSLS-Probit	50	0.02	-2.24	4.21	2.70	10.99	83.2
		FVR	50	0.09	0.52	1.30	0.05	307.29	1.3
		FBVR	50	0.03	0.61	1.33	0.03	374.96	0.6
2	$\theta = 0.3$	OLS	50	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.49	1.08	1.89	0.28	112.08	0.0
		TSLS	50	0.02	-0.47	2.51	0.58	23.15	65.9
		TSLS-Probit	50	0.02	-0.46	2.49	0.57	23.57	66
		FVR	50	0.09	0.65	1.17	0.03	694.54	0.0
		FBVR	50	0.03	0.74	1.21	0.02	843.08	0.1
3	$\theta = 0.4$	OLS	50	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.49	1.18	1.80	0.26	206.71	0
		TSLS	50	0.02	-0.06	2.09	0.30	41.03	50.1
		TSLS-Probit	50	0.02	-0.04	2.07	0.29	42.30	49.5
		FVR	50	0.09	0.72	1.11	0.02	1,257.50	0.0
		FBVR	50	0.03	0.8	1.15	0.01	1,527.31	0.0
4	$\theta = 0.5$	OLS	50	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.49	1.24	1.74	0.26	338.66	0.0
		TSLS	50	0.01	0.17	1.86	0.19	63.24	35.2
		TSLS-Probit	50	0.01	0.19	1.84	0.18	66.29	33.4
		FVR	50	0.09	0.76	1.07	0.01	2,007.95	0.0
		FBVR	50	0.03	0.83	1.12	0.01	2,436.82	0.0
5	$\theta = 0.6$	OLS	50	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.49	1.27	1.71	0.25	519.87	0.0
		TSLS	50	0.01	0.31	1.72	0.13	90.80	19.5
		TSLS-Probit	50	0.01	0.33	1.69	0.12	97.08	16.8
		FVR	50	0.09	0.78	1.04	0.01	2,975.80	0.0
		FBVR	50	0.03	0.85	1.09	0.00	3,616.12	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathcal{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both *i.i.d.* $N(0,1)$. Each numbered group of rows corresponds to a unique empirical setting where only the strength of the instrument θ varies. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\rho = 0.4$, $\delta = 50$, $\mu = -1.663$, the exclusion restriction $\mathcal{Z}_{it}$ is normally distributed, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$ to keep the share of units treated equal to 50% in all settings. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldridge's modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 7: main simulation results when endogeneity ρ is higher

Empirical setting	Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
6 $\theta = 0.2$ $\rho = 0.6$	OLS	50	0.26	1.20	1.32	0.07		0.0
	TSLTE	50	0.73	1.14	2.32	0.63	50	0.6
	TSLS	50	0.04	-2.87	4.94	3.98	11	83.0
	TSLS-Probit	50	0.04	-3.09	5.16	4.43	11	82.9
	FVR	50	0.14	0.48	1.25	0.06	310	1.7
	FBVR	50	0.04	0.59	1.33	0.04	381	0.7
7 $\theta = 0.3$ $\rho = 0.6$	OLS	50	0.26	1.20	1.32	0.07		0.0
	TSLTE	50	0.73	1.34	2.12	0.58	113	0.0
	TSLS	50	0.01	-0.48	2.51	0.58	23	65.3
	TSLS-Probit	50	0.01	-0.47	2.49	0.57	24	64.8
	FVR	50	0.14	0.60	1.12	0.04	700	0.0
	FBVR	50	0.04	0.71	1.20	0.02	849	0.0
8 $\theta = 0.4$ $\rho = 0.6$	OLS	50	0.26	1.21	1.32	0.07		0.0
	TSLTE	50	0.73	1.43	2.03	0.56	207	0.0
	TSLS	50	0.02	-0.06	2.09	0.30	41	50.4
	TSLS-Probit	50	0.01	-0.04	2.07	0.29	42	49.8
	FVR	50	0.13	0.67	1.06	0.03	1,257	0.0
	FBVR	50	0.04	0.78	1.14	0.01	1,525	0.0
9 $\theta = 0.5$ $\rho = 0.6$	OLS	50	0.26	1.21	1.31	0.07		0.0
	TSLTE	50	0.74	1.49	1.98	0.56	338	0.0
	TSLS	50	0.01	0.17	1.86	0.19	63	34.8
	TSLS-Probit	50	0.01	0.19	1.84	0.18	67	33.4
	FVR	50	0.14	0.71	1.02	0.02	2,007	0.0
	FBVR	50	0.04	0.81	1.11	0.01	2,433	0.0
10 $\theta = 0.6$ $\rho = 0.6$	OLS	50	0.26	1.21	1.31	0.07		0.0
	TSLTE	50	0.74	1.53	1.95	0.56	521	0.0
	TSLS	50	0.01	0.31	1.71	0.13	91	18.7
	TSLS-Probit	50	0.01	0.34	1.68	0.12	97	15.6
	FVR	50	0.14	0.73	0.99	0.02	2,972	0.0
	FBVR	50	0.04	0.84	1.08	0.01	3,617	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both i.i.d.$N(0,1)$. Each numbered group of rows corresponds to a unique empirical setting where only the strength of the instrument θ varies. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\rho = 0.6$, $\delta = 50$, $\mu = -1.663$, the exclusion restriction $\mathscr{Z}_i$ is normally distributed, and the parameter λ equals $\sqrt{(1 - \theta^2 - \overline{\rho^2})}$ to keep the share of units treated equal to 50% in all settings. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldridge's modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 8: simulation results with different T

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
1	$T = 4$	OLS	50	0.36	1.26	1.46	0.13		0.0
		TSLTE	50	0.14	0.69	1.58	0.07	112	1.1
		TSLS	50	0.01	0.26	1.71	0.14	56	24.8
		TSLS-Probit	50	0.02	0.27	1.70	0.13	58	23.9
		FVR	50	0.10	0.51	1.30	0.05	186	1.5
		FBVR	50	0.02	0.63	1.33	0.03	238	0.2
2	$T = 8$	OLS	50	0.26	1.19	1.34	0.07		0.0
		TSLTE	50	0.34	0.98	1.70	0.15	165	0.0
		TSLS	50	0.00	0.15	1.85	0.19	52	35.2
		TSLS-Probit	50	0.00	0.17	1.83	0.18	54	34.4
		FVR	50	0.11	0.63	1.15	0.03	547	0.0
		FBVR	50	0.03	0.73	1.21	0.02	674	0.0
3	$T = 12$	OLS	50	0.20	1.14	1.26	0.04		0.0
		TSLTE	50	0.44	1.10	1.78	0.23	193	0.0
		TSLS	50	0.00	0.00	2.01	0.26	46	43.2
		TSLS-Probit	50	0.00	0.01	1.99	0.26	47	42.6
		FVR	50	0.10	0.68	1.11	0.02	945	0.0
		FBVR	50	0.03	0.77	1.16	0.01	1,157	0.0
4	$T = 16$	OLS	50	0.17	1.11	1.22	0.03		0.0
		TSLTE	50	0.50	1.20	1.80	0.28	211	0.0
		TSLS	50	0.00	-0.10	2.11	0.32	40	51.3
		TSLS-Probit	50	0.00	-0.09	2.09	0.31	41	49.8
		FVR	50	0.09	0.72	1.10	0.02	1,375	0.0
		FBVR	50	0.03	0.80	1.15	0.01	1,669	0.0
5	$T = 20$	OLS	50	0.14	1.09	1.19	0.02		0.0
		TSLTE	50	0.55	1.24	1.85	0.32	228	0.0
		TSLS	50	0.00	-0.17	2.18	0.36	35	58.0
		TSLS-Probit	50	0.00	-0.15	2.16	0.35	36	57.0
		FVR	50	0.08	0.76	1.08	0.01	1,834	0.0
		FBVR	50	0.02	0.83	1.12	0.01	2,224	0.0
6	$T = 24$	OLS	50	0.12	1.08	1.16	0.02		0.0
		TSLTE	50	0.59	1.30	1.89	0.37	240	0.0
		TSLS	50	0.01	-0.23	2.21	0.39	32	59.3
		TSLS-Probit	50	0.01	-0.22	2.19	0.38	33	57.9
		FVR	50	0.07	0.79	1.07	0.01	2,301	0.0
		FBVR	50	0.02	0.85	1.11	0.00	2,803	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both *i.i.d.*$N(0,1)$. Each numbered group of rows corresponds to a unique empirical setting where only T and μ varies. We change μ to keep the share of units treated close to 50% for each setting to make them comparable. In all setting, $n = 1,000$, $\alpha = \beta = 1$, $\theta = 0.4$, $\rho = 0.4$, $\delta = 50$, the exclusion restriction $\mathscr{Z}_{it}$ is normally distributed, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$ to keep the share of units treated equal to 50% in all settings. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldridge's modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 9: simulation results with different T when endogeneity ρ is higher

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
7	$T = 4$	OLS	50	0.54	1.44	1.64	0.29		0.0
		TSLTE	50	0.20	0.76	1.65	0.09	112	0.8
		TSLS	50	0.02	0.25	1.71	0.14	56	24.6
		TSLS-Probit	50	0.02	0.26	1.70	0.13	58	23.7
		FVR	50	0.14	0.47	1.25	0.06	187	3.2
		FBVR	50	0.03	0.62	1.32	0.03	239	0.1
8	$T = 8$	OLS	50	0.39	1.32	1.46	0.16		0.0
		TSLTE	50	0.51	1.16	1.85	0.29	164	0.0
		TSLS	50	0.00	0.15	1.85	0.19	53	36.4
		TSLS-Probit	50	0.00	0.16	1.83	0.18	55	35.5
		FVR	50	0.17	0.57	1.10	0.05	547	0.0
		FBVR	50	0.05	0.70	1.19	0.02	673	0.0
9	$T = 12$	OLS	50	0.30	1.25	1.36	0.09		0.0
		TSLTE	50	0.66	1.34	1.98	0.46	193	0.0
		TSLS	50	0.00	-0.01	2.01	0.27	46	42.6
		TSLS-Probit	50	0.00	0.01	2.00	0.26	47	42.2
		FVR	50	0.15	0.63	1.06	0.04	943	0.0
		FBVR	50	0.05	0.75	1.15	0.01	1156	0.0
10	$T = 16$	OLS	50	0.25	1.19	1.30	0.06		0.0
		TSLTE	50	0.76	1.46	2.06	0.60	212	0.0
		TSLS	50	0.00	-0.11	2.10	0.32	40	51.2
		TSLS-Probit	50	0.00	-0.09	2.08	0.31	41	50.4
		FVR	50	0.13	0.68	1.06	0.03	1374	0.0
		FBVR	50	0.04	0.78	1.14	0.01	1673	0.0
11	$T = 20$	OLS	50	0.21	1.17	1.26	0.05		0.0
		TSLTE	50	0.83	1.53	2.12	0.70	227	0.0
		TSLS	50	0.00	-0.18	2.18	0.36	35	56.6
		TSLS-Probit	50	0.00	-0.16	2.16	0.35	36	55.8
		FVR	50	0.12	0.72	1.04	0.02	1828	0.0
		FBVR	50	0.04	0.81	1.11	0.01	2217	0.0
12	$T = 24$	OLS	50	0.18	1.14	1.23	0.03		0.0
		TSLTE	50	0.89	1.61	2.17	0.81	239	0.0
		TSLS	50	0.01	-0.24	2.21	0.39	32	60.0
		TSLS-Probit	50	0.01	-0.22	2.19	0.38	33	59.3
		FVR	50	0.11	0.75	1.04	0.02	2293	0.0
		FBVR	50	0.03	0.83	1.10	0.01	2787	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both *i.i.d.* $N(0,1)$. Each numbered group of rows corresponds to a unique empirical setting where only T and μ varies. We change μ to keep the share of units treated close to 50% for each setting to make them comparable. In all setting, $n = 1,000$, $\alpha = \beta = 1$, $\theta = 0.4$, $\rho = 0.6$, $\delta = 50$, the exclusion restriction $\mathscr{Z}_{it}$ is normally distributed, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$ to keep the share of units treated equal to 50% in all settings. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldridge's modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 10: variable definitions and descriptive statistics

Variable	Description	Level of aggrega-tion	No. obs.	Mean	Std. Dev.	Min	Max
Unavailability factor (UF)	100 - (Actual operating hours / Potential operating hours) x 100	Reactor	1,851	12.397	15.132	0	100
Divest (0/1)	Equals 1 if the reactor is divested. 0 otherwise	Reactor	2,369	0.204	0.403	0	1
Age	Age of the reactor (in years)	Reactor	1,851	23.733	8.299	1	43
Ind	Share of state level electricity consumption from industrial consumers	State	2,266	21.77	7.707	4.974	48.25
Rep	Equals 1 if both state senate and house of representatives have a Republican majority	State	2,266	0.299	0.458	0	1

Table 11: estimation output of model (8.1)

	OLS	TSLS	TSLS-Probit	FVR	FBVR
$Divest(0/1)$	-7.143***	7.095	5.762	-6.771	-7.624***
	(1.452)	(13.460)	(6.976)	(5.594)	(2.202)
Age	0.150	-0.216	-0.182	0.141	0.162
	(0.218)	(0.419)	(0.243)	(0.252)	(0.210)
Age^2	-0.010**	-0.010**	-0.010**	-0.010**	-0.010**
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Year dummies	Yes	Yes	Yes	Yes	Yes
Treatment of obs. where UF = 100	Dummy variable	Dummy variable	Dummy variable	Dummy variable	Dummy variable
R^2	0.39	0.37	0.38	0.44	0.44
No. obs.	1,851	1,851	1,851	1,851	1,851

Notes. The dependent variable is UF, the total number of outage hours divided by maximum potential generation hours. For all column except OLS, the share of state level electricity consumption by industrial customers in the previous period is the unique original instrument for the divestiture treatment. SE in brackets are robust to heteroskedasticity and autocorrelation with a Bartlett bandwidth = 2. *$p < 0.10$, **$p < 0.05$, *** $p < 0.01$.

Appendices

A Proofs of propositions

Proof of lemma 3.1. It holds

$$
\begin{aligned}
P\{T_i = k\} &= P\{-\theta Z_{i1} + U_{i1} \ge 0 \ldots, -\theta Z_{ik-1} + U_{ik-1} \ge 0, -\theta Z_{ik} + U_{ik} < 0\} \\
&= P\{-\theta Z_{i1} + U_{i1} \ge 0\} \ldots P\{-\theta Z_{ik-1} + U_{ik-1} \ge 0\} P\{-\theta Z_{ik} + U_{ik} < 0\} \\
&= (\frac{1}{2})^k, \hspace{5cm} \square
\end{aligned}
$$

where the second equality is due to the independence assumption and the third due to $-\theta Z_{ik-1} + U_{ik-1} \sim N(0,1)$. Thus we obtain

$$
P\{T_i \le t\} = \sum_{k=1}^{t} P\{T_i = k\} = \sum_{k=1}^{t} (\frac{1}{2})^k = 1 - (\frac{1}{2})^t \to 1.
$$

with $t \to \infty$. Because we have assumed $D_{it} = 1$ if $D_{it-1} = 1$, condition (PT1) holds as well.

Proof of proposition 3.0.1. Let P denote the probability measure, with respect to which all one dimensional variables $(Y_{it}, D_{it}, U_{it}, Z_{it})$ are measurable.
  First, we have

$$
\begin{aligned}
\mathbb{E}[D_{it}] &= \int D_{it} dP = \int \mathbb{1}\{D_{it} = 1\} D_{it} dP + \int \mathbb{1}\{D_{it} = 0\} D_{it} dP \\
&= \int \mathbb{1}\{D_{it} = 1\}.1 dP = P\{D_{it} = 1\} \xrightarrow{t \to \infty} 1. \hspace{2cm} \square
\end{aligned}
$$

For the covariance $cov(D_{it}, Z_{it})$, we have

$$
\begin{aligned}
\mid cov(D_{it}, Z_{it}) \mid &= \mid \mathbb{E}[D_{it} Z_{it}] - \mathbb{E}[D_{it}] \mathbb{E}[Z_{it}] \mid = \mid \int D_{it} Z_{it} dP - \int D_{it} dP \int Z_{it} dP \mid \\
&\le \mid \int D_{it} Z_{it} dP - \int Z_{it} dP \mid + \mid \int Z_{it} dP - \int D_{it} dP \int Z_{it} dP \mid \\
&=: A + B.
\end{aligned}
$$

For A, we obtain

$$
\begin{aligned}
A &= \mid \int (D_{it} - 1) Z_{it} dP \mid = \mid \int \mathbb{1}\{D_{it} = 1\}(D_{it} - 1) Z_{it} dP + \int \mathbb{1}\{D_{it} = 0\}(D_{it} - 1) Z_{it} dP \mid \\
&= \mid \int \mathbb{1}\{D_{it} = 0\} Z_{it} dP \mid .
\end{aligned}
$$

Since per assumption $\mathbb{1}\{D_{it} = 0\} = o_P(1)$ and $Z_{it} = O_P(1)$, it holds that $\mathbb{1}\{D_{it} = 0\}Z_{it} = o_P(1)$ and therefore $A = o(1)$.

For B, we obtain

$$B = |\int Z_{it}dP(1 - \int D_{it}dP)| = O(1)o(1) = o(1).$$

As a result, $cov(D_{it}, Z_{it}) = o(1)$.

Proof of proposition (3.0.2). We observe first that $D_{it} = \mathbb{1}\{T_i < t\} + \mathbb{1}\{T_i \geq t\}\mathbb{1}\{U_{it} < \theta Z_{it}\}$. Using this equality, we obtain

$$\begin{aligned}
\mathbb{E}[D_{it}Z_{it}] &= \mathbb{E}[(\mathbb{1}\{T_i < t\} + \mathbb{1}\{T_i \geq t\}\mathbb{1}\{U_{it} < \theta Z_{it}\})Z_{it}] \\
&= \mathbb{E}[\mathbb{1}\{T_i < t\}Z_{it}] + \mathbb{E}[\mathbb{1}\{T_i \geq t\}\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}] \\
&=: A + B
\end{aligned}$$
$\square$

Since $\mathbb{1}\{T_i < t\} = \sum_{k=1}^{t-1} \mathbb{1}\{T_i = k\}$, it holds that

$$A = \sum_{k=1}^{t-1} \mathbb{E}[\mathbb{1}\{T_i = k\}Z_{it}] = \sum_{k=1}^{t-1} \mathbb{E}[\mathbb{1}\{T_i = k\}]\mathbb{E}[Z_{it}] = \sum_{k=1}^{t-1} \mathbb{E}[\mathbb{1}\{T_i = k\}]0 = 0,$$

where the second equality holds due to independence of the instrument and the error and the third equality holds per assumption.

Further, we observe that $\mathbb{1}\{T_i < t\} = \sum_{k=1}^{t-1} \mathbb{1}\{T_i = k\}$ depends only on $U_{i1}, \ldots U_{it-1}$ and on $Z_{i1}, \ldots Z_{it-1}$. Consequently, $\mathbb{1}\{T_i < t\}$ is independent from U_{it} and Z_{it} per assumption and therefore independent from $\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}$. Using this independence and writing $\mathbb{1}\{T_i < t\} + \mathbb{1}\{T_i \geq t\} = 1$, we obtain for B

$$\begin{aligned}
B &= \mathbb{E}[\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}] - \mathbb{E}[\mathbb{1}\{T_i < t\}\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}] \\
&= \mathbb{E}[\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}] - \mathbb{E}[\mathbb{1}\{T_i < t\}]\mathbb{E}[\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}] \\
&= \mathbb{E}[\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}](1 - P\{T_i < t\}) = C(1 - P\{T_i < t\}) = C\left(\frac{1}{2}\right)^t,
\end{aligned}$$

where C is a constant equal to $\mathbb{E}[\mathbb{1}\{U_{it} < \theta Z_{it}\}Z_{it}]$ which per construction does not depend on t. It follows that

$$\frac{\sum_{t=1}^{T} \mathbb{E}[D_{it}Z_{it}]}{\sqrt{T}} = \frac{C\sum_{t=1}^{T} 2^{-t}}{\sqrt{T}} = o(1), \tag{A.1}$$

which completes the proof.

Proof of proposition 4.1. We provide a proof for the case of a single (endogenous) covariate and a single instrument. The generalisation to the case of additional covariates and multiple instruments follows the same steps and is omitted. Assume first that $\bar{T} = T$. It holds

$$\mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip}] = 0 \tag{A.2}$$

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

$$\mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip}] = \mathbb{E}[\mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip} \mid T_i]]$$

$$= \sum_{t=1}^{T} \mathbb{E}[\mathbb{1}\{t = k\}Z_{il}U_{ip} \mid T_i = t]P\{T_i = t\} = 0$$

$$+ \mathbb{E}[Z_{il}U_{ip} \mid T_i = k]P\{T_i = k\} = 0. \qquad \square$$

The last equality holds due to assumption R1 and

$$\mathbb{E}[Z_{il}U_{ip} \mid T_i = k] = \mathbb{E}[\mathbb{E}[Z_{il}U_{ip} \mid Z_{il}, T_i = k] \mid T_i = k]$$
$$= \mathbb{E}[Z_{il}\mathbb{E}[U_{ip} \mid Z_{il}, T_i = k] \mid T_i = k]$$

Due to analogical arguments, it holds

$$\mathbb{E}[T_i Z_{il}U_{ip}] = 0 \tag{A.3}$$

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

Identification of FVR: Multiply model 4.3 with $\widetilde{\widetilde{Z}}_i$ to obtain

$$\widetilde{\widetilde{Z}}_i'\widetilde{Y}_i = \widetilde{\widetilde{Z}}_i'\widetilde{W}_i\gamma + \widetilde{\widetilde{Z}}_i'\widetilde{U}_i. \tag{A.4}$$

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

$$Z_{ik}\mathbb{1}\{T_i \geq k\} + Z_{iT_i}\mathbb{1}\{T_i < k\} - T^{-1}\sum_{l=1}^{T_i} Z_{il} - T^{-1}(T - T_i)Z_{iT_i} \tag{A.5}$$

We prove now 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.2 and because $\mathbb{1}\{T_i \geq k\} = \sum_{l=k}^{T} \mathbb{1}\{T_i = l\}$. For the second term, observe that $Z_{iT_i} = \sum_{l=i}^{T} \mathbb{1}\{T_i = l\}Z_{il}$ and $\mathbb{1}\{T_i = l\}\mathbb{1}\{T_i = m\} = 0$ whenever $l \neq m$ and then analogous argument as for the first term applies. For the third term, observe that $\sum_{l=1}^{T_i} Z_{il} = \sum_{k=1}^{T} \mathbb{1}\{T_i = k\}\sum_{l=1}^{k} Z_{il}$ and hence $\mathbb{E}[\sum_{l=1}^{T_i} Z_{il}U_{ip}] = \sum_{k=1}^{T}\sum_{l=1}^{k} \mathbb{E}[\mathbb{1}\{T_i = k\}Z_{il}U_{ip}] = 0$. Finally,

$\mathbb{E}[Z_{il}U_{ip}] = 0$ and with A.3 we obtain $\mathbb{E}[T^{-1}(T - T_i)Z_{iT_i}Z_{il}U_{ip}] = 0$. Therefore, all summands in the sum $\widetilde{Z}'_i \widetilde{U}_i$ have an expectation zero. Thus $\mathbb{E}[\widetilde{Z}'_i \widetilde{U}_i] = 0$. Using assumptions R2 a) and R3 a), we finally obtain

$$\gamma = \mathbb{E}[\Pi\widetilde{Z}'_i \widetilde{W}_i]^{-1} \mathbb{E}[\Pi\widetilde{Z}'_i \widetilde{Y}_i], \tag{A.6}$$

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

Identification of FBVR: The proof follows identical steps as for the FVR approach. Observe that the k-th element of $\widetilde{\overset{\circ}{Z}}_i$ 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 $\check{Z}_{ik} = \mathbb{1}\{T_i > k\}Z_{iT_i-1} + \mathbb{1}\{T_i \le k\}Z_{iT_i}$. Therefore, showing that $\mathbb{E}[\widetilde{\overset{\circ}{Z}}'_i \widetilde{U}_i] = 0$ amounts to showing that

$$\mathbb{E}[\mathbb{1}\{T_i > k\}Z_{iT_i-1}U_{it}] - \mathbb{E}[\mathbb{1}\{T_i > k\}Z_{iT_i}U_{it}] + \mathbb{E}[T_iZ_{iT_i-1}U_{it}] - \mathbb{E}[T_iZ_{iT_i}U_{it}] = 0, \tag{A.7}$$

which was already established above.

The case of $\bar{T} > T$ goes analogically by introducing $\mathbb{1}\{T_i > T\}$ and $\mathbb{1}\{T_i \le T\}$. In the first case, all expressions are the same as in the standard TSLS case, and the second case is as above.

B Monte Carlo Simulations: robustness checks

Distribution of the original instrument

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 12 and Table 13, we report simulations where $\mathscr{Z}$ 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. The only difference between the tables is the endogeneity level ρ that is higher in Table 13 than in Table 12. These results are similar to those shown in Table 6 and Table 7 where FVR and FBVR are found to be the most efficient estimators. There are two differences. First, the variances of the competing estimators FVR, FBVR, TSLS and TSLS-Probit are smaller than when $\mathscr{Z}$ is drawn from a normal distribution. Second, the efficiency gap equal to a factor 2.5 is lower than in the normal distribution case but remain large.

Treatment frequency

In our baseline results, the share of treated units equals 50%. In Table 14 and Table 15, we perform simulations in which the share of treated units varies from 30% to 90%.[19] The only difference between the tables is the endogeneity level ρ that is higher in Table 15 than in Table 14. Between 50% and 90% the efficiency of FVR and FBVR does not vary significantly and both estimators do not generate type II errors. At 30%, the confidence intervals of the FVR and FBVR estimators are slightly wider, whereas the TSLS and TSLS-Probit generates a significant amount of type II errors. Finally the efficiency gap, measured by the ratio between 95% confidence interval of TSLS and the interval of F(B)VR, is increasing the share of units treated. This is because the dataset used for F(B)VR and the dataset used for TSLS become more similar as the share of treated units reduces.[20]

---

[19] We start at 30% because estimations with TSLS and TSLS-Probit do not converge at lower shares.

[20] The similarity comes from the fact that at low share of units treated the transformed instrument is applied to a lower proportion of units.

Table 12: simulation results with χ^2-distributed original instrument $\mathscr{L}$

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
1	$\theta = 0.2$	OLS	48	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	48	0.22	0.66	1.78	0.13	60	2.6
		TSLS	48	0.02	-0.48	2.45	0.56	25	66.7
		TSLS-Probit	48	0.01	-0.46	2.43	0.54	26	65.6
		FVR	48	0.07	0.67	1.20	0.02	647	0.1
		FBVR	48	0.01	0.74	1.24	0.02	757	0.0
2	$\theta = 0.3$	OLS	49	0.17	1.11	1.22	0.03		0.0
	$\rho = 0.4$	TSLTE	49	0.11	0.74	1.48	0.05	142	0.0
		TSLS	49	0.00	0.20	1.79	0.17	72	30.9
		TSLS-Probit	49	0.00	0.21	1.78	0.16	74	29.6
		FVR	49	0.05	0.78	1.11	0.01	1,838	0.0
		FBVR	49	0.01	0.84	1.14	0.01	2,109	0.0
3	$\theta = 0.4$	OLS	50	0.16	1.10	1.21	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.04	0.77	1.31	0.02	276	0.0
		TSLS	50	0.00	0.46	1.54	0.08	148	5.8
		TSLS-Probit	50	0.00	0.47	1.53	0.07	153	5.4
		FVR	50	0.04	0.84	1.08	0.01	3,887	0.0
		FBVR	50	0.00	0.89	1.11	0.00	4,429	0.0
4	$\theta = 0.5$	OLS	51	0.14	1.09	1.19	0.02		0.0
	$\rho = 0.4$	TSLTE	51	0.01	0.79	1.24	0.01	488	0.0
		TSLS	51	0.00	0.59	1.41	0.04	249	0.5
		TSLS-Probit	51	0.00	0.59	1.41	0.04	258	0.4
		FVR	51	0.04	0.87	1.06	0.00	7,027	0.0
		FBVR	51	0.00	0.91	1.08	0.00	8,019	0.0
5	$\theta = 0.6$	OLS	52	0.13	1.07	1.18	0.02		0.0
	$\rho = 0.4$	TSLTE	52	0.02	0.83	1.21	0.01	814	0.0
		TSLS	52	0.00	0.66	1.34	0.03	362	0.0
		TSLS-Probit	52	0.00	0.67	1.33	0.03	375	0.0
		FVR	52	0.03	0.89	1.05	0.00	11,439	0.0
		FBVR	52	0.00	0.92	1.07	0.00	13,151	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{L}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both i.i.d.$N(0, 1)$. Each numbered group of rows corresponds to a unique empirical setting where only the strength of the instrument θ varies. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\rho = 0.4$, $\delta = 50$, $\mu = -1.689$, the exclusion restriction $\mathscr{L}_{it}$ is drawn from a standardized χ^2-distribution, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$ to keep the share of units treated equal to 50% in all settings. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldrige modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 13: simulation results with χ^2-distributed original instrument $\mathscr{Z}$ when endogeneity ρ is higher

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
6	$\theta = 0.2$	OLS	48	0.26	1.20	1.31	0.07		0.0
	$\rho = 0.6$	TSLITE	48	0.34	0.79	1.88	0.19	60	2.0
		TSLS	48	0.02	-0.51	2.46	0.57	25	66.6
		TSLS-Probit	48	0.02	-0.49	2.44	0.56	26	65.5
		FVR	48	0.10	0.63	1.17	0.03	647	0.1
		FBVR	48	0.03	0.72	1.23	0.02	754	0.0
7	$\theta = 0.3$	OLS	49	0.25	1.19	1.30	0.06		0.0
	$\rho = 0.6$	TSLITE	49	0.16	0.80	1.52	0.06	143	0.0
		TSLS	49	0.01	0.20	1.79	0.16	72	30.7
		TSLS-Probit	49	0.00	0.21	1.78	0.16	75	29.2
		FVR	49	0.08	0.75	1.08	0.01	1,840	0.0
		FBVR	49	0.02	0.83	1.14	0.01	2,115	0.0
8	$\theta = 0.4$	OLS	50	0.23	1.18	1.29	0.06		0.0
	$\rho = 0.6$	TSLITE	50	0.06	0.79	1.33	0.02	277	0.0
		TSLS	50	0.00	0.46	1.54	0.08	149	6.0
		TSLS-Probit	50	0.00	0.47	1.53	0.07	154	5.4
		FVR	50	0.07	0.81	1.05	0.01	3,899	0.0
		FBVR	50	0.01	0.88	1.10	0.00	4,445	0.0
9	$\theta = 0.5$	OLS	51	0.21	1.16	1.27	0.05		0.0
	$\rho = 0.6$	TSLITE	51	0.02	0.79	1.25	0.01	487	0.0
		TSLS	51	0.00	0.59	1.41	0.04	249	0.4
		TSLS-Probit	51	0.00	0.59	1.41	0.04	258	0.4
		FVR	51	0.06	0.85	1.04	0.01	7,024	0.0
		FBVR	51	0.00	0.91	1.08	0.00	8,018	0.0
10	$\theta = 0.6$	OLS	52	0.19	1.14	1.24	0.04		0.0
	$\rho = 0.6$	TSLITE	52	0.03	0.84	1.22	0.01	813	0.0
		TSLS	52	0.00	0.66	1.34	0.03	362	0.0
		TSLS-Probit	52	0.00	0.67	1.33	0.03	374	0.0
		FVR	52	0.05	0.87	1.03	0.00	11,444	0.0
		FBVR	52	0.00	0.92	1.07	0.00	13,167	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both *i.i.d.* $N(0, 1)$. Each numbered group of rows corresponds to a unique empirical setting where only the strength of the instrument θ varies. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\rho = 0.6$, $\delta = 50$, $\mu = -1.689$, the exclusion restriction $\mathscr{Z}_{it}$ is drawn from a standardized χ^2-distribution, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$ to keep the share of units treated equal to 50% in all settings. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldrige modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 14: simulation results with different shares of units treated

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
1	$\theta = 0.4$	OLS	30	0.19	1.11	1.26	0.04		0.0
	$\rho = 0.4$	TSLTE	30	0.62	1.26	1.98	0.42	150	0.0
		TSLS	30	0.02	-0.71	2.75	0.78	28	72.9
		TSLS-Probit	30	0.02	-0.66	2.69	0.73	30	71.3
		FVR	30	0.12	0.58	1.19	0.04	826	0.1
		FBVR	30	0.08	0.63	1.21	0.03	923	0.1
2	$\theta = 0.4$	OLS	50	0.17	1.12	1.23	0.03		0.0
	$\rho = 0.4$	TSLTE	50	0.49	1.18	1.80	0.26	207	0.0
		TSLS	50	0.02	-0.06	2.09	0.30	41	50.1
		TSLS-Probit	50	0.02	-0.04	2.07	0.29	42	49.5
		FVR	50	0.09	0.72	1.11	0.02	1,258	0.0
		FBVR	50	0.03	0.80	1.15	0.01	1,527	0.0
3	$\theta = 0.4$	OLS	60	0.17	1.12	1.22	0.03		0.0
	$\rho = 0.4$	TSLTE	60	0.43	1.13	1.73	0.21	228	0.0
		TSLS	60	0.02	0.10	1.93	0.22	46	39.7
		TSLS-Probit	60	0.01	0.10	1.92	0.22	47	39.7
		FVR	60	0.07	0.76	1.09	0.01	1,467	0.0
		FBVR	60	0.00	0.86	1.14	0.01	1,862	0.0
4	$\theta = 0.4$	OLS	75	0.17	1.12	1.21	0.03		0.0
	$\rho = 0.4$	TSLTE	75	0.33	1.05	1.61	0.13	249	0.0
		TSLS	75	0.01	0.26	1.77	0.15	54	25.0
		TSLS-Probit	75	0.01	0.26	1.77	0.15	55	25.0
		FVR	75	0.05	0.81	1.08	0.01	1,792	0.0
		FBVR	75	0.04	0.92	1.15	0.00	2,433	0.0
5	$\theta = 0.4$	OLS	90	0.16	1.12	1.20	0.03		0.0
	$\rho = 0.4$	TSLTE	90	0.22	0.93	1.51	0.07	251	0.0
		TSLS	90	0.01	0.36	1.67	0.11	62	14.1
		TSLS-Probit	90	0.01	0.36	1.67	0.11	62	14.5
		FVR	90	0.02	0.86	1.10	0.00	2,117	0.0
		FBVR	90	0.08	0.98	1.18	0.01	3,061	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both $i.i.d.N(0,1)$. Each numbered group of rows corresponds to a unique empirical setting where only *mu* varies in order to choose the share of units treated. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\theta = 0.4$, $\rho = 0.4$, $\delta = 50$, the exclusion restriction $\mathscr{Z}_{it}$ is normally distributed, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldrige modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

Table 15: simulation results with different shares of units treated when endogeneity ρ is higher

Empirical setting		Estimator	Average share of units treated (%)	Absolute bias	95% lower bound	95% upper bound	RMSE	Average first stage F-stat	Type II error frequency (%)
6	$\theta = 0.4$	OLS	30	0.28	1.20	1.35	0.08		0.0
	$\rho = 0.6$	TSLTE	30	0.94	1.58	2.30	0.92	150	0.0
		TSLS	30	0.01	-0.73	2.76	0.79	28	72.9
		TSLS-Probit	30	0.01	-0.68	2.70	0.75	30	71.5
		FVR	30	0.18	0.52	1.13	0.06	827	0.2
		FBVR	30	0.12	0.58	1.18	0.04	924	0.2
7	$\theta = 0.4$	OLS	50	0.26	1.21	1.32	0.07		0.0
	$\rho = 0.6$	TSLTE	50	0.73	1.43	2.03	0.56	207	0.0
		TSLS	50	0.02	-0.06	2.09	0.30	41	50.4
		TSLS-Probit	50	0.01	-0.04	2.07	0.29	42	49.8
		FVR	50	0.13	0.67	1.06	0.03	1,257	0.0
		FBVR	50	0.04	0.78	1.14	0.01	1,525	0.0
8	$\theta = 0.4$	OLS	60	0.25	1.20	1.30	0.07		0.0
	$\rho = 0.6$	TSLTE	60	0.65	1.36	1.93	0.44	228	0.0
		TSLS	60	0.01	0.11	1.92	0.22	47	39.8
		TSLS-Probit	60	0.01	0.11	1.91	0.21	48	39.3
		FVR	60	0.11	0.72	1.05	0.02	1,473	0.0
		FBVR	60	0.00	0.85	1.15	0.01	1,869	0.0
9	$\theta = 0.4$	OLS	75	0.25	1.20	1.29	0.06		0.0
	$\rho = 0.6$	TSLTE	75	0.50	1.23	1.78	0.27	250	0.0
		TSLS	75	0.01	0.26	1.76	0.15	55	25.7
		TSLS-Probit	75	0.01	0.26	1.76	0.15	55	25.6
		FVR	75	0.08	0.78	1.05	0.01	1,794	0.0
		FBVR	75	0.05	0.94	1.17	0.01	2,441	0.0
10	$\theta = 0.4$	OLS	90	0.25	1.21	1.29	0.06		0.0
	$\rho = 0.6$	TSLTE	90	0.33	1.04	1.62	0.13	251	0.0
		TSLS	90	0.01	0.35	1.67	0.11	62	15.3
		TSLS-Probit	90	0.01	0.35	1.67	0.11	62	15.5
		FVR	90	0.04	0.85	1.08	0.00	2,115	0.0
		FBVR	90	0.12	1.03	1.22	0.02	3,065	0.0

Notes. Simulation results from 1,000 replications generated from the estimation of parameter α of the structural model $Y_{it} = \alpha D_{it} + X_{it}\beta + C_i + U_{it}$ generated from the discrete choice model $D_{it} = \mathbb{1}\{\mu + \delta D_{it-1} + \theta \mathscr{Z}_{it} + X_{it}\beta + \rho U_{it} + \epsilon_{it} > 0\}$. 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 U_{it} and the error term ϵ_{it} of the discrete choice model are both $i.i.d. N(0,1)$. Each numbered group of rows corresponds to a unique empirical setting where only mu varies in order to choose the share of units treated. In all setting, $n = 1,000$, $T = 15$, $\alpha = \beta = 1$, $\theta = 0.4$, $\rho = 0.6$, $\delta = 50$, the exclusion restriction $\mathscr{Z}_{it}$ is normally distributed, and the parameter λ equals $\sqrt{(1 - \theta^2 - \rho^2)}$. TSLS denotes the two stage least square estimates, TSLTE denotes the two stage local treatment estimator which corresponds to a TSLS estimator using only the observations before and after the treatment time, TSLS-Probit is Wooldrige modification of TSLS, FVR is the Forward Variation Reduction estimator, and FBVR is the Forward and Backward Variation Reduction estimator.

C   Test of identifying assumption

Table 16: size of the test statistics, number of rejections out of 100 cases

t = 1		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	0.09	0.121	-0.084	0.014	0.129	-0.13	-0.035	-0.133	-0.035	0.017
	Variance	0.826	1.136	1.159	1.018	1.195	0.795	0.83	1.347	0.996	0.97
	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	1.11	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	1.16	0.83	1.075	1.324
	10%	0	0	0	0	0	0	0	0	0	0
p = 6	Mean	-0.15	0.045	-0.028	-0.014	-0.154	-0.078	0.05	0.09	0.032	0.016
	Variance	1.249	1.136	0.889	0.98	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	1.21	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	0.13	-0.078	0.081	0.097
	Variance	0.979	0.921	0.994	0.974	1.049	0.963	0.979	1.037	1.17	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	-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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 17: size of the test statistics, number of rejections out of 100 cases

t = 2		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	0.057	-0.13	-0.106	0.149	0.078	0.069	0.081	0.013	0.042	-0.12
	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	-0.05	-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	0.9
	10%	0	0	0	0	0	0	0	0	0	0
p = 4	Mean	-0.047	-0.131	-0.237	-0.16	0.077	-0.013	0.028	0.014	-0.041	-0.015
	Variance	1.043	0.95	1.101	1.316	1.453	0.808	0.923	0.948	0.98	1.02
	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	-0.07	-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	1.28
	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	-0.09	0.123
	Variance	1.162	1.045	0.939	1.189	1.014	0.942	0.891	1.198	1.022	1.09
	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	0.13	-0.226	-0.153	-0.135	0.022	-0.06	0.142	0.154	0.02	-0.016
	Variance	1.055	0.887	1.229	1.294	0.754	1.232	1.08	1.066	1.218	1.013
	10%	0	0	0	0	0	0	0	0	0	0
p = 11	Mean	0.041	0.262	0.03	0.145	-0.045	-0.172	-0.049	-0.004	-0.057	0.156
	Variance	0.845	1.076	0.99	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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1', a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 18: size of the test statistics, number of rejections out of 100 cases

t = 3		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	-0.088	-0.125	-0.027	0.07	-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	-0.09	-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	1.19	1.293	0.895	1.102	0.969
	10%	0	0	0	0	0	0	0	0	0	0
p = 4	Mean	0.08	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	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	0.11	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	1.16	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	-0.13	0.247	-0.025	-0.035	-0.465	-0.08	0.13	0.135	-0.122
	Variance	1.162	1.249	0.925	1.27	0.942	0.936	0.967	1	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	-0.19	-0.082	0.11	-0.135
	Variance	1.101	1.172	1.189	0.929	1.101	0.85	1.078	1.042	1.023	1.413
	10%	0	0	0	0	0	0	0	0	0	0
p = 9	Mean	-0.13	0.025	-0.056	-0.192	0.064	0.04	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	0.03	-0.068	0.025	0.026	0.098	0.059	0.049	-0.145	0.104
	Variance	0.927	1.421	1	0.911	0.999	1.434	0.752	1.178	1.015	0.93
	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	0.11	-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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 19: size of the test statistics, number of rejections out of 100 cases

t = 4		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	0.084	0.207	0.083	-0.063	-0.114	-0.006	0.19	0.018	0.061	-0.159
	Variance	1.014	1.127	1.169	0.997	1.109	1.08	0.981	1.097	1.042	0.84
	10%	0	0	0	0	0	0	0	0	0	0
p = 2	Mean	-0.038	-0.16	-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	0.02	-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	1.17	0.938
	10%	0	0	0	0	0	0	0	0	0	0
p = 4	Mean	0.09	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	-0.19	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	1.19	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	-0.01	0.118	-0.205	-0.184	-0.038	0.187	-0.021
	Variance	0.905	0.831	0.778	1.093	0.99	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	-0.1
	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	0.02	0.009	-0.071	-0.158	0.082
	Variance	0.937	0.897	1	1.086	1.165	0.814	0.939	1.177	1.171	1.29
	10%	0	0	0	0	0	0	0	0	0	0
p = 10	Mean	-0.018	-0.052	0.117	0.044	-0.036	-0.05	0.076	0.008	-0.172	0.196
	Variance	0.78	0.962	0.957	1.07	1.24	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	-0.02	0.119	-0.084	-0.03	0.099	-0.016	0.111	-0.005
	Variance	0.968	1.164	0.98	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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathcal{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathcal{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 20: size of the test statistics, number of rejections out of 100 cases

t = 5		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	0.121	0.146	0.018	-0.169	0.068	-0.077	-0.002	0.05	-0.061	-0.054
	Variance	1.084	1.12	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	-0.08
	Variance	0.777	1.001	1.24	0.929	1.203	1.43	1.035	1.089	1.101	1.014
	10%	0	0	0	0	0	0	0	0	0	0
p = 3	Mean	0.128	0.21	0.014	-0.344	-0.152	0.021	0.06	0.013	0.024	0.047
	Variance	0.889	1.11	1.139	0.91	1.219	1.065	1.202	0.97	0.958	1.144
	10%	0	0	0	0	0	0	0	0	0	0
p = 4	Mean	-0.18	-0.004	0.122	-0.325	0.066	-0.059	0.081	-0.006	0.14	0.093
	Variance	1.226	0.905	1.152	1.257	1.04	1.197	1.075	0.974	0.886	0.981
	10%	0	0	0	0	0	0	0	0	0	0
p = 5	Mean	0.04	-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	1.14	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	0.08	0.155	0.072	0.013	0.002	0.085	0.109	0.056	0.158
	Variance	1.422	0.9	1.066	1.09	0.859	0.961	1.262	0.763	1.02	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	-0.01	-0.35	-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	-0.11	0.096	-0.014	0.007	0.033	-0.251	-0.125	-0.177
	Variance	1.05	0.986	1.327	1.157	0.93	1.051	1.176	1.28	1.03	0.971
	10%	0	0	0	0	0	0	0	0	0	0
p = 10	Mean	0.081	0.04	-0.092	-0.037	-0.202	0.134	-0.098	-0.123	-0.062	-0.14
	Variance	1.605	0.986	1.173	1.18	1.166	1.194	0.93	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	-0.02	-0.239
	Variance	1.028	1.295	0.92	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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathcal{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathcal{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 21: size of the test statistics, number of rejections out of 100 cases

t = 6		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	-0.065	-0.061	0.131	0.098	-0.077	-0.009	0.03	0.069	0.022	-0.023
	Variance	0.99	0.899	1.062	1.178	0.931	0.91	0.931	1.131	1.118	0.828
	10%	0	0	0	0	0	0	0	0	0	0
p = 2	Mean	0	-0.018	0.094	-0.01	-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	1.02
	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	-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	-0.02	0.143
	Variance	1.085	0.882	1.095	1.16	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	-0.06	0.023	0.093	-0.021	2.66	-0.182	-0.178	-0.083	-0.151	-0.089
	Variance	0.966	1.25	1.024	0.739	0.7	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	-0.04	0.173	0.01	0.106	-0.126	0.077	0.05	0.164	-0.003
	Variance	0.931	0.918	1.267	1.078	1.083	1.171	0.8	1.044	1.15	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	1.21	0.89
	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	-0.1
	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	-0.03	-0.036
	Variance	1.053	0.66	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	0.2	0.201	0.034	0.157	-0.225
	Variance	0.754	1.209	0.96	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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 22: size of the test statistics, number of rejections out of 100 cases

t = 7		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	-0.102	-0.162	-0.089	0.093	-0.13	-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	1.07
	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	0.96	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	1.14	1.17	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	-0.24	-0.061	0.148	0.007	-0.103
	Variance	0.936	1.057	0.949	1.082	1.224	1.108	0.922	1.02	0.906	1.287
	10%	0	0	0	0	0	0	0	0	0	0
p = 5	Mean	-0.017	-0.062	-0.08	0.016	0.038	-0.15	-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	1.14	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	-0.08	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	1.22	0.782	1.267	0.727	1.16	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	0.04	0.088	0	-0.1	-0.099	-0.273	0.026
	Variance	1.094	0.89	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	-0.2	0.04	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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathcal{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathcal{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 23: size of the test statistics, number of rejections out of 100 cases

t = 8		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
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	1.05	1.04	0.996	1.09	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	-0.21	-0.142	-0.012	0.046
	Variance	1.111	0.812	0.906	0.8	0.868	0.974	0.968	1.062	1.018	1.11
	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	1.18	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	0.29	-0.042	-0.077	0.077	-0.028	-0.124	-0.268	-0.149	0.016	-0.096
	Variance	1.114	1.12	1.033	1.035	1.1	1.359	1.019	0.98	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	1.1	0.856	0.955	0.945	0.844	1.096	1.144	0.877	1.312	1.18
	10%	0	0	0	0	0	0	0	0	0	0
p = 7	Mean	-0.09	-0.056	-0.113	-0.027	0.177	0.268	-0.09	-0.082	0.145	-0.024
	Variance	1.037	1.06	1.086	1.152	1.061	1.082	1.071	1.291	1.15	0.91
	10%	0	0	0	0	0	0	0	0	0	0
p = 8	Mean	-0.024	0.15	-0.177	-0.02	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	0.81	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	-0.07
	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	-0.06	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	0.82	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	0.14	-0.099	-0.167	-0.003	-0.03	-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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1', a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 24: size of the test statistics, number of rejections out of 100 cases

t = 9		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	-0.118	-0.016	0.063	0.096	0.06	-0.004	-0.027	0.085	-0.025	0.243
	Variance	1.388	0.968	1.097	0.974	1.119	0.927	0.93	0.994	1.031	0.954
	10%	0	0	0	0	0	0	0	0	0	0
p = 2	Mean	-0.166	-0.149	0.18	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	1.21	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	-0.08	-0.147	0.176	0.041
	Variance	1.055	0.891	1.137	0.993	0.819	1.174	1.24	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	0.04	-0.098
	Variance	1.168	1.143	1.001	1.036	1.079	1.034	1.181	1.103	1.328	0.95
	10%	0	0	0	0	0	0	0	0	0	0
p = 6	Mean	-0.134	0.14	-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	-0.35	0.135	0.111
	Variance	1.276	1.005	0.901	0.887	1.04	0.987	1.16	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	1.07	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	1.24	1.01	0.907
	10%	0	0	0	0	0	0	0	0	0	0
p = 10	Mean	0.021	-0.082	0.17	0.158	0.003	-0.168	-0.004	0.137	-0.184	0.275
	Variance	1.11	1.058	1.345	1.122	1.075	0.984	1.091	1.204	0.85	1.015
	10%	0	0	0	0	0	0	0	0	0	0
p = 11	Mean	0.069	0.081	0.053	0.08	-0.011	0.066	-0.064	-0.02	-0.062	-0.071
	Variance	1.089	0.719	1.213	1.14	0.804	1.235	1.217	1.176	1.069	0.853
	10%	0	0	0	0	0	0	0	0	0	0

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathcal{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathcal{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 25: size of the test statistics, number of rejections out of 100 cases

t = 10		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	0.105	0.173	0	-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	0.11
	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	-0.09	-0.033	-0.002	-0.201	-0.008	-0.016
	Variance	1.013	0.813	0.946	0.919	1.134	1.217	1.08	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	0.02	-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	-0.2
	Variance	1.011	1.122	1.304	1.052	1.676	1.208	1.073	0.78	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	0.88	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	1.05	0.87	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	0.98
	10%	0	0	0	0	0	0	0	0	0	0
p = 9	Mean	0.218	-0.112	-0.102	0.234	-0.1	-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	0.08	0.102	2.478	-0.074
	Variance	1.058	1.16	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	0.1	-0.19	0.152	-0.233	0.09	-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

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1′, a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

Table 26: size of the test statistics, number of rejections out of 100 cases

t = 11		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10	k = 11
p = 1	Mean	0.069	-0.11	0.003	-0.087	0.047	-0.018	0.009	0.033	-0.02	0.035
	Variance	1.115	1.135	1.004	1.06	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	-0.02	-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	0.19	0.078	0.014	-0.133
	Variance	0.889	1.116	0.875	0.896	1.02	0.943	1.072	0.95	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	1.16
	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	0.04	0.008	0.045	0.087	-0.161	0.08	0.077	0.036	-0.121
	Variance	1.119	1.031	0.851	1.215	1.198	0.969	1.153	0.974	1.11	0.944
	10%	0	0	0	0	0	0	0	0	0	0
p = 7	Mean	-0.17	-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	-0.11	0.004	-0.088	-0.035	-0.089
	Variance	0.886	0.963	1.278	1.063	0.724	1.264	1.22	0.941	1.115	1.19
	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	0.05	0.121	-0.081	0.068	-0.174	0.117	-0.052
	Variance	1.098	0.96	0.978	0.978	0.857	1.09	1.14	0.881	1.161	1.28
	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	0.92	1.112	0.994	1.165
	10%	0	0	0	0	0	0	0	0	0	0

Notes. The data generating process is the same as in previous sections except that we generate only 100 samples and $T = 11$ for tractability. We set $\theta = 0.4$, $\rho = 0.4$, and $\mu = -1.662$. For each t, k, p we calculate the sample average $\bar{P}_{ktp} = \frac{1}{n_k} \sum_{i \in \mathcal{K}} \mathscr{Z}_{it} U_{ip}$ and the sample standard deviation $s_{ktp} = \sqrt{\frac{1}{n_k - 1} \sum_{i \in K} (\mathscr{Z}_{it} U_{ip} - \bar{P}_{ktp})^2}$ where n_k is the number of units treated in period k. Under R1', a hypothesis that holds under R1 that is the identifying assumption of the F(B)VR estimator, the statistic $\mathcal{T}_{ktp} = \frac{\bar{P}_{ktp}}{s_{ktp}/\sqrt{n_k}}$ has a Student distribution $t(n_k - 1)$.

D   Estimation with an alternative instrument

A natural test of robustness is to use an alternative instrument. In Table 27, we replicate the estimation approaches presented in Table 11 using a dummy of a state Republican majority as an instrumental variable. We expect this dummy variable to influence the divestiture decision since Republican politicians are reportedly more favourable 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 11. 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 27: estimation output of model (8.1) when an indicator for state Republican majority is used as instrument

	TSLS	TSLS-Probit	FVR	FBVR
Divest(0/1)	-17.630**	-7.567	-19.125***	-10.959**
	(6.338)	(5.380)	(5.540)	(4.364)
Age	0.420	0.161	0.458*	0.248
	(0.277)	(0.229)	(0.263)	(0.228)
Age^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	Dummy variable	Dummy variable	Dummy variable	Dummy variable
R^2	0.40	0.44	0.39	0.43
No. obs.	1,851	1,851	1,851	1,851

Notes. The dependent variable is UF, the total number of outage hours divided by maximum potential generation hours. For all columns except OLS, a dummy equal to 1 when the state has a Republican majority in the previous period is the unique original instrument for the divestiture treatment. SE in brackets are robust to heteroskedasticity and autocorrelation with a Bartlett bandwidth = 2. *$p < 0.10$, **$p < 0.05$, *** $p < 0.01$.

E   Spillover

Results in Section 8 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 labour movements.[21] 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 11 using the share of industrial electricity consumption as an instrument.

We present the results in Table 28. 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) and (iii) for nuclear reactors located in the same state (column 3) or in the same and neighbouring states (column 4). We add the appropriate dummy variables to our base specification for each scenario.[22]

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.

---

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

[22]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 20122013. Available at: www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1350/appa.xls

Table 28: estimation output of model (8.1) allowing for spillovers

	Operational Spillovers	Technical Spillovers	Geographic Spillovers	
	(1)	(2)	(3)	(4)
$Divest(0/1)$	-7.716***	-7.189 **	-8.368 **	-9.447
	(2.266)	(3.280)	(3.958)	(6.596)
$SameOperator(0/1) \times Divest(0/1)$	-1.275			
	(1.596)			
$SameTechnology(0/1) \times Divest(0/1)$		-0.720		
		(2.126)		
$SameLocation(0/1) \times Divest(0/1)$			-0.862	-1.834
			(2.313)	(4.7)
Age	0.169	0.144	0.200	0.259
	(0.210)	(0.344)	(0.253)	(0.361)
Age^2	-0.009**	-0.010***	-0.009**	-0.009**
	(0.003)	(0.004)	(0.004)	(0.004)
Year dummies	Yes	Yes	Yes	Yes
Treatment of obs. where UF = 100	Dummy variable	Dummy variable	Dummy variable	Dummy variable
R^2	0.44	0.44	0.40	0.43
No. obs.	1,851	1,851	1,851	1,851

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. The dependent variable is UF, the total number of outage hours divided by maximum potential generation hours. For all columns except OLS, the share of state level electricity consumption by industrial customers in the previous period is the unique original instrument for the divestiture treatment. SE in brackets are robust to heteroskedasticity and autocorrelation with a Bartlett bandwidth = 2. *$p < 0.10$, **$p < 0.05$, *** $p < 0.01$.