Sufficient Instruments Filter for Causal Discovery

Rajveer Jat

University of California, Riverside

October 21, 2024

Introduction

- *Endogeneity*
  - Makes OLS estimates biased.
  - Instrumental Variable (IV) technique to isolate exogenous variation in x_t.

$$
\begin{aligned}
x_t =& \alpha_0 + \alpha_1 z_t + e_t \\
y_t =& \beta_0 + \beta_1 x_t + \varepsilon_t
\end{aligned}
$$

Introduction

- *Endogeneity*
  - Makes OLS estimates biased.
  - Instrumental Variable (IV) technique to isolate exogenous variation in x_t.

$$x_t = \alpha_0 + \alpha_1 z_t + e_t$$
$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

- If z is an instrument:
  - Why not z^2, z^3...,z^{100}, e^z, $z^2 \cos^{-1}(\sin(\frac{\pi}{2} z))$? and their combinations.
    - Ignoring these forms $\sim$ Model Mis-specification $\sim$ Invalid Analysis.
    - How to use all of them? Use Non-parametric IV (NPIV).

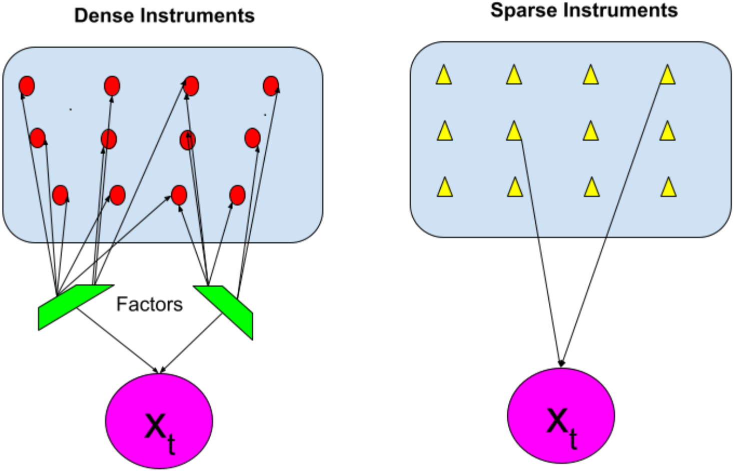
Introduction

- *Endogeneity*
  - Makes OLS estimates biased.
  - Instrumental Variable (IV) technique to isolate exogenous variation in x_t.

$$x_t = \alpha_0 + \alpha_1 z_t + e_t$$
$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

- If z is an instrument:
  - Why not z^2, $z^3 \ldots, z^{100}$, e^z, $z^2 \cos^{-1}(\sin(\frac{\pi}{2} z))$? and their combinations.
    - Ignoring these forms $\sim$ Model Mis-specification $\sim$ Invalid Analysis.
    - How to use all of them? Use Non-parametric IV (NPIV).
- NPIV may be too slow when the number of z is sizeable.
  - Too slow requires many data points and computational resources.
  - How to address this problem then?
    - Put some structure on the set of instruments.
    - How? (Next Page)

Structure on Many Instruments



Example: Price Elasticity of Automobile Demand

Berry, Levinsohn, and Pake (ECTA, 1995) (hereafter BLP) estimates α_0:

$$y_{it} = \alpha_0 p_{it} + x_{it}'\beta_0 + \varepsilon_{it} \quad (1)$$
$$p_{it} = z_{it}'\delta_0 + x_{it}'\gamma_0 + u_{it} \quad (2)$$

- Many Instruments:
  - Product characteristics of the competitors.

Example: Price Elasticity of Automobile Demand

Berry, Levinsohn, and Pake (ECTA, 1995) (hereafter BLP) estimates α_0:

$$y_{it} = \alpha_0 p_{it} + x_{it}'\beta_0 + \varepsilon_{it} \quad (1)$$

$$p_{it} = z_{it}'\delta_0 + x_{it}'\gamma_0 + u_{it} \quad (2)$$

- Many Instruments:
  - Product characteristics of the competitors.
- Notes that 2SLS yields positive α.
  - Because the 2SLS cannot handle many instruments.
  - Their estimate is -0.145 which makes sense.

Example: Price Elasticity of Automobile Demand

Berry, Levinsohn, and Pake (ECTA, 1995) (hereafter BLP) estimates α_0:

$$y_{it} = \alpha_0 p_{it} + x_{it}'\beta_0 + \varepsilon_{it} \quad (1)$$
$$p_{it} = z_{it}'\delta_0 + x_{it}'\gamma_0 + u_{it} \quad (2)$$

- Many Instruments:
  - Product characteristics of the competitors.
- Notes that 2SLS yields positive α.
  - Because the 2SLS cannot handle many instruments.
  - Their estimate is -0.145 which makes sense.
- Chernozhukov et al (AER, 2015):
  - Considers many functional forms of BLP's instruments.
  - Assumes a sparse structure on instruments.
  - Finds more negative α_0 than BLP.

Structure on Many Instruments

- **Sparsity**
  - Only a few out of many instruments are relevant.
  - Belloni et al (*Econometrica*, 2012): Post-LASSO IV (PLIV).
    - Selection of the most relevant instruments will ideal approach.
    - Allows non-linearities through sieves i.e. includes z^2, z^3,.. etc as regressors.
    - Either non-linearity (through sieves) or number of z being large is allowed.

Structure on Many Instruments

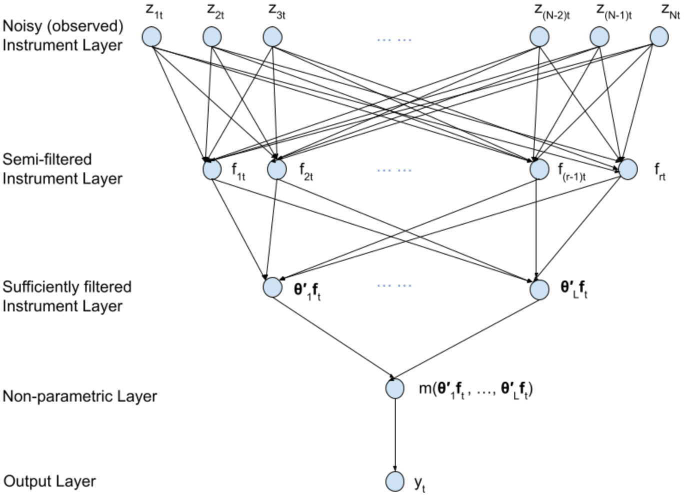
- **Sparsity**
  - Only a few out of many instruments are relevant.
  - Belloni et al (*Econometrica*, 2012): Post-LASSO IV (PLIV).
    - Selection of the most relevant instruments will ideal approach.
    - Allows non-linearities through sieves i.e. includes z^2, z^3,.. etc as regressors.
    - Either non-linearity (through sieves) or number of z being large is allowed.
- **Factor Structure:**
  - Many instruments share common components.
  - Bai and Ng (*Econometric Theory, 2012*): Factor Instrumental Variables (FIV).
    - Common components (factors) can be used as instruments.
    - Only number of z being large is allowed.
    - Can't handle Non-linearities.

Structure on Many Instruments

- **Sparsity**
  - Only a few out of many instruments are relevant.
  - Belloni et al (*Econometrica*, 2012): Post-LASSO IV (PLIV).
    - Selection of the most relevant instruments will ideal approach.
    - Allows non-linearities through sieves i.e. includes z^2, z^3,.. etc as regressors.
    - Either non-linearity (through sieves) or number of z being large is allowed.
- **Factor Structure:**
  - Many instruments share common components.
  - Bai and Ng (*Econometric Theory, 2012*): Factor Instrumental Variables (FIV).
    - Common components (factors) can be used as instruments.
    - Only number of z being large is allowed.
    - Can't handle Non-linearities.
- Both method fails when: non-linearities + the number of z being large.
- *Why not 2SLS?*
  - 2SLS is inconsistent with many instruments [Bekker (ECTA, 1994), Berry et al (ECTA, 1995)].

This Paper

- Filters sufficient information from many instruments for efficient estimation.



This Paper: Attractive Features

- **Requires Weaker Assumptions**
  - Assumes that at least one linear combination of factors is a valid instrument.
    - Some of z may be weak or even invalid instruments.
    - Some of the common factors may be weak instruments.
- **Supervision**
  - Some of the factors of set of z may not be relevant for x.
    - SIF filters them out, making the estimation procedure more efficient.
- **Handles Non-Linearities**
  - We estimate the "first-stage" $(\widehat{x})$ as non-parametric function of instruments.
    - Alleviates mis-specification problem.
- **Sufficient Dimension Reduction**
  - Reduces the number of dimensions by encapsulating the information.
    - Can handle the case: Number of Instruments $(N) >$ Sample Size (T)

Why is This Paper Important?

Let's see how Belloni et al (2012) perform under dense instruments. $[\beta_1 = 2]$

				$E(\widehat{\beta_1})$		$\text{RMSE}(\widehat{\beta}_1)$	
Design	ρ	T	N	OLS	Belloni et al	OLS	Belloni et al
Linear	0	200	50	2.06	2.05	0.11	0.11
Linear	0	200	150	2.06	2.05	0.11	0.11
Linear	0	200	250	2.06	2.05	0.11	0.11
Linear	0.5	200	50	2.33	2.93	0.35	0.94
Linear	0.5	200	150	2.33	2.82	0.35	0.83
Linear	0.5	200	250	2.33	2.73	0.35	0.75
Linear	0.9	200	50	2.61	3.05	0.62	1.06
Linear	0.9	200	150	2.61	2.96	0.62	0.97
Linear	0.9	200	250	2.61	2.92	0.62	0.93
Non-linear	0	200	50	2.05	2.06	0.12	0.14
Non-linear	0	200	150	2.05	2.07	0.12	0.14
Non-linear	0	200	250	2.05	2.06	0.12	0.13
Non-linear	0.5	200	50	2.25	3.07	0.28	1.10
Non-linear	0.5	200	150	2.25	2.92	0.28	0.96
Non-linear	0.5	200	250	2.25	2.82	0.28	0.87
Non-linear	0.9	200	50	2.45	3.21	0.46	1.24
Non-linear	0.9	200	150	2.45	3.07	0.46	1.10
Non-linear	0.9	200	250	2.45	2.98	0.46	1.02

Revisiting BLP Example

- Denote Z=Matrix of instruments used in Chernozhukov et al (2015).
  - They use Belloni et al (2012)'s Sparsity-based method.
- Let's see if their instruments are sparse.

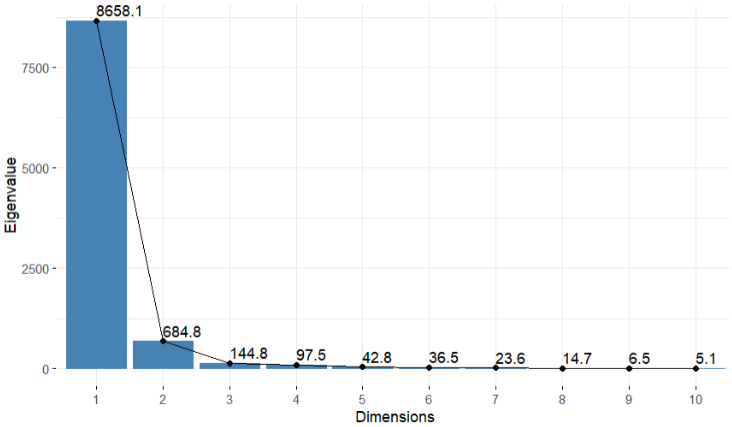


Figure: Scree-plot of Eigenvalues of Z'Z

Our Framework

At a given time t:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (3)$$

$$x_t = m\left(\boldsymbol{\theta}_1' \mathbf{f}_t, \ldots, \boldsymbol{\theta}_L' \mathbf{f}_t\right) + e_t \quad (4)$$

$$z_{it} = \mathbf{b}_i' \mathbf{f}_t + u_{it}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq T \quad (5)$$

$$f_{jt} = \gamma_j f_{jt-1} + v_{jt}, \quad 1 \leq j \leq r \quad (6)$$

- y_t is target, x_t is endogenous regressor, $\boldsymbol{z}_i$ are observed "noisy" instruments.
- $\mathbf{f}_t$ is vector of r factors. u_{it} and v_{it} are weakly autocorrelated errors.
- Errors (ε_t, e_t) are allowed to be auto-correlated and cross correlated.

Our Framework

At a given time t:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{3}$$

$$x_t = m\left(\boldsymbol{\theta}_1' \mathbf{f}_t, \ldots, \boldsymbol{\theta}_L' \mathbf{f}_t\right) + e_t \tag{4}$$

$$z_{it} = \mathbf{b}_i' \mathbf{f}_t + u_{it}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq T \tag{5}$$

$$f_{jt} = \gamma_j f_{jt-1} + v_{jt}, \quad 1 \leq j \leq r \tag{6}$$

- y_t is target, x_t is endogenous regressor, $\boldsymbol{z}_i$ are observed "noisy" instruments.
- $\mathbf{f}_t$ is vector of r factors. u_{it} and v_{it} are weakly autocorrelated errors.
- Errors (ε_t, e_t) are allowed to be auto-correlated and cross correlated.
- $(\boldsymbol{\theta}_1' \mathbf{f}_t, \ldots, \boldsymbol{\theta}_L' \mathbf{f}_t)$ are called Sufficient Dimension Reduction (SDR) indices.
- $\boldsymbol{b}_i$: factor loadings. $1 \leq L \leq r \leq N$.
- N: number of instruments, T: sample-size. $N > T$ is allowed.

Estimation of Factors

- We use Principal Components to estimate factors.
  - Ref: Stock and Watson (JASA, 2002), Bai (ECTA, 2003).

$$\left(\widehat{\mathbf{B}}_r, \widehat{\mathbf{F}}_r\right) = \arg\min_{(\mathbf{B},\mathbf{F})} \left\|\mathbf{Z} - \mathbf{B}\mathbf{F}'\right\|_F^2 \tag{7}$$

$$\text{subject to } T^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_r, \quad \mathbf{B}'\mathbf{B} \text{ is diagonal} \tag{8}$$

Where

- $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T), \mathbf{F}' = (\mathbf{f}_1, \ldots, \mathbf{f}_T)$
- $\|\cdot\|_F$ denotes the Frobenius norm.
  - Square root of the sum of the absolute squares of its matrix elements.

Introduction to SDRs

- $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L)$ are Sufficient Dimension Reduction (SDR) directions if:

$$x_t \perp \mathbf{f_t} \mid (\boldsymbol{\theta}_1' \mathbf{f}_t, \ldots, \boldsymbol{\theta}_L' \mathbf{f}_t)$$

Introduction to SDRs

- $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L)$ are Sufficient Dimension Reduction (SDR) directions if:

$$x_t \perp \mathbf{f_t} \mid (\boldsymbol{\theta}_1' \mathbf{f}_t, \dots, \boldsymbol{\theta}_L' \mathbf{f}_t)$$

- $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L)$ are the orthonormal basis of Central Subspace $S_{x|\mathbf{f}_t}$.
  - $(\boldsymbol{\theta}_1' \mathbf{f}_t, \dots, \boldsymbol{\theta}_L' \mathbf{f}_t)$ are called SDR indices for x_t.
  - If x_t is linearly related with factors $\implies L = 1$.

Introduction to SDRs

- $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L)$ are Sufficient Dimension Reduction (SDR) directions if:

$$x_t \perp \mathbf{f_t} \mid (\boldsymbol{\theta}_1' \mathbf{f}_t, \ldots, \boldsymbol{\theta}_L' \mathbf{f}_t)$$

- $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L)$ are the orthonormal basis of Central Subspace $S_{x|\mathbf{f}_t}$.
    - $(\boldsymbol{\theta}_1' \mathbf{f}_t, \ldots, \boldsymbol{\theta}_L' \mathbf{f}_t)$ are called SDR indices for x_t.
    - If x_t is linearly related with factors $\implies L = 1$.

- How to obtain Central Subspace $S_{x|\mathbf{f}_t}$:
    - Li (JASA, 1991) says $S_{x|\mathbf{f}_t}$ contains the linear span of $\text{cov}\left(E\left(\mathbf{f}_t \mid x_t\right)\right)$.
        - $E\left(\mathbf{f}_t \mid x_t\right)$ is inverse regression. Non-parametrically estimated.
        - Inverse regression is the source of supervision in the method.

- Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L)$ and $E\left(\mathbf{f}_t \mid x_t\right) = \Theta \mathbf{a}\left(x_t\right)$. Then,

$$\text{cov}\left(E\left(\mathbf{f}_t \mid x_t\right)\right) = \Theta E\left[\mathbf{a}\left(x_t\right)\mathbf{a}\left(x_t\right)^T\right]\Theta'$$

Estimation of SDRs

- Let's order x_t and divide it into M slices $(I_1, \ldots, I_M)$. Then,

$$\begin{aligned}\Sigma_{\mathbf{f}|x} =& \frac{1}{M}\sum_{s=1}^{M} E\left(\mathbf{f}_t \mid x_t \in I_s\right) E\left(\mathbf{f}_t' \mid x_t \in I_s\right) \\ =& \boldsymbol{\Theta}\left[\frac{1}{M}\sum_{s=1}^{M} E\left(\mathbf{a}\left(x_t\right) \mid x_t \in I_s\right) E\left(\mathbf{a}\left(x_t\right) \mid x_t \in I_s\right)^T\right]\boldsymbol{\Theta}'\end{aligned}$$

Estimation of SDRs

- Let's order x_t and divide it into M slices $(I_1, \ldots, I_M)$. Then,

$$\begin{aligned}\Sigma_{\mathbf{f}|x} =& \frac{1}{M}\sum_{s=1}^{M} E\left(\mathbf{f}_t \mid x_t \in I_s\right) E\left(\mathbf{f}_t' \mid x_t \in I_s\right) \\ =& \boldsymbol{\Theta}\left[\frac{1}{M}\sum_{s=1}^{M} E\left(\mathbf{a}\left(x_t\right) \mid x_t \in I_s\right) E\left(\mathbf{a}\left(x_t\right) \mid x_t \in I_s\right)^T\right]\boldsymbol{\Theta}'\end{aligned}$$

- Slicing:
$\left\{\left(x_{(s,j)}, \widehat{\mathbf{f}}_{(s,j)}\right) : x_{(s,j)} = x_{(c(s-1)+j+1)},\ \widehat{\mathbf{f}}_{(s,j)} = \widehat{\mathbf{f}}_{(c(s-1)+j)}\right\}_{s=1,\ldots,M;j=1,\ldots,c}$.

- The Estimator of Central Subspace $S_{x|\mathbf{f}_t}$:

$$\widehat{\Sigma}_{\mathbf{f}|x} = \frac{1}{M}\sum_{s=1}^{M}\left[\frac{1}{c}\sum_{j=1}^{c}\widehat{\mathbf{f}}_{(s,j)}\right]\left[\frac{1}{c}\sum_{j=1}^{c}\widehat{\mathbf{f}}_{(s,j)}\right]'. \tag{9}$$

- SDR directions are first L eigenvectors of $\widehat{\Sigma}_{\mathbf{f}|x}$.

Estimation of Non-parametric Function

- For simplicity, call $\boldsymbol{\theta_j}'\mathbf{f} = \boldsymbol{w}_j$, we want to estimate $m(\boldsymbol{w}_j, \ldots, \boldsymbol{w}_L)$.
  - Definition: $m(\boldsymbol{w}) = \int x_t g(x_t \mid \boldsymbol{w}) dx_t$.
    - $g(x_t \mid \boldsymbol{w})$ is conditional density.
  - Fan and Gijbels (AoS, 1992)'s LLLS used for estimation of $m(\cdot)$.
- Bandwidth
  - We use the same bandwidth for all arguments for simplicity.
  - Optimal bandwidth is chosen by cross-validation

$$h_{opt} \propto T^{-1/(L+4)}$$

- Kernel:
  - Gaussian symmetric and integrating to one kernel is used.
  - More structure on the kernel function is in Assumption-5.

Estimation of β

- **First-Stage**: Obtain $\widehat{x}_t = \widehat{m}(\widehat{\boldsymbol{w}}_t)$ from previous steps.
  - Note that $\widehat{x}_t$ is completely determined by $\widehat{\boldsymbol{w}}_t = \hat{\boldsymbol{\theta}}'\widehat{\mathbf{f}}_t$.
    - Factors $\widehat{\mathbf{f}}_t$ are exogenous source of variation.
- **Second-Stage**: Replace x by $\widehat{x}$ in the equation for y.
  - Use OLS to estimate β.

Algorithm

Algorithm 1	Sufficient Instrument Filter (SIF) Procedure
Step 1	Obtain the estimated factors $\left\{\widehat{\mathbf{f}}_t\right\}_{t=1,\ldots,T}$ from 7 and 8.
Step 2	Construct $\widehat{\Sigma}_{\mathbf{f}\|x}$ described in the equation-9.
Step 3	Obtain $\widehat{\boldsymbol{\theta}}_1, \cdots, \widehat{\boldsymbol{\theta}}_L$ from the L largest eigenvectors of $\widehat{\Sigma}_{\mathbf{f}\|x}$.
Step 4	Construct the predictive indices $\widehat{\boldsymbol{\theta}}_1'\widehat{\mathbf{f}}_t, \cdots, \widehat{\boldsymbol{\theta}}_L'\widehat{\mathbf{f}}_t$.
Step 5	Use the local linear least squared regression to estimate $m(\cdot)$ with indices from Step 4, and hence to get $\widehat{x}_t$.
Step 6	Use the $\widehat{x}_t$ obtained in step-5 in place of x_t in equation-3 and do OLS to get $\widehat{\beta} = (\hat{\beta}_0 \quad \hat{\beta}_1)'$

Tuning Parameters

- Number of Factors r
  - We use the eigenvalue ratio test by Ahn and Horenstei (ECTA, 2013).
- Number of Slices M
  - Number of slices does not matter much.
  - Fan et al (JoE, 2017) suggests $M \geq \max\{L, 2\}$ is good enough.
    - They use $M = 10$, we follow their guide and use same.
- Number of SDRs L
  - The first L eigenvalues of $\widehat{\Sigma}_{\mathbf{f}|x}$ must be significantly different from zero compared to the estimation error.
  - Methods for determining L: Li (JASA, 1991) and Schott (JASA, 1994).

Asymptotic Theory: The Plan

1. Assumptions.
2. Identification of β.
3. Consistency of Intermediate Steps Estimation.
4. Consistency of β.
5. Asymptotic Normality of β.
    - Sketch of Proof.

Assumptions-1: Identification Assumptions

Assumption

1. $E\left[m(\boldsymbol{\theta_1}'\mathbf{f}_t, \ldots, \boldsymbol{\theta_L}'\mathbf{f}_t)x_t\right] \neq 0$ *Relevancy Condition*
2. $E\left[m(\boldsymbol{\theta_1}'\mathbf{f}_t, \ldots, \boldsymbol{\theta_L}'\mathbf{f}_t)\varepsilon_t\right] = 0$ *Exclusion Restriction*
3. $E\left[m(\boldsymbol{\theta_1}'\mathbf{f}_t, \ldots, \boldsymbol{\theta_L}'\mathbf{f}_t)e_t\right] = 0$ *Model Assumption*
4. $\mathbb{E}[\varepsilon_t] = 0$, *and* $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t - E(\varepsilon_t)\right) = O_p(1)$ *for all* t.
5. $\mathbb{E}[e_t] = 0$, *and* $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}e_t - E(e_t)\right) = O_p(1)$ *for all* t.

Assumptions-2: Factors, Loadings, SDR

Assumption

1. **Pervasive Condition:** *The loadings* $\mathbf{b}_i$ *satisfy* $\|\mathbf{b}_i\| \leq \mathcal{M}$ *for* $i = 1, \ldots, N$. *As* $N \to \infty$, *there exist two positive constants* c_1 *and* c_2 *such that:*

$$c_1 < \lambda_{\min}\left(\frac{1}{N}\mathbf{B}'\mathbf{B}\right) < \lambda_{\max}\left(\frac{1}{N}\mathbf{B}'\mathbf{B}\right) < c_2$$

2. **Identification:** $\frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{I}_K$, *and* $\mathbf{B}'\mathbf{B}$ *is a diagonal matrix with distinct entries.*

3. **Linearity:** *The expectation* $E\left(\mathbf{b}'\mathbf{f}_t \mid \boldsymbol{\theta}_1'\mathbf{f}_t, \ldots, \boldsymbol{\theta}_L'\mathbf{f}_t\right)$ *is a linear function of* $\boldsymbol{\theta}_1'\mathbf{f}_t, \ldots, \boldsymbol{\theta}_L'\mathbf{f}_t$ *for any* $\mathbf{b} \in \mathbb{R}^N$, *where the vectors* $\boldsymbol{\theta}_i'$ *derived from model 4.*

- This is the same as Assumption-3.1 of Fan et al (JoE, 2017)

Assumptions-3: Data Generating Process

Assumption

$\{\mathbf{f}_t\}_{t\geq 1}$, $\{\mathbf{u}_t\}_{t\geq 1}$, and $\{e_t\}_{t\geq 1}$ are strictly stationary processes and mutually independent. Additionally, $E\left\|\mathbf{f}_t\right\|^4 < \infty$ and $E\left(\left\|\mathbf{f}_t\right\|^2 \mid x_t\right) < \infty$. For some positice constant c, the mixing coefficient $\alpha(T) < c\rho^T$ for all $T \in \mathbb{Z}^+$ and some $\rho \in (0,1)$.

- Same as Assumption 3.1 of Fan et al (JoE, 2017).
- Same as Assumption A(d) in Bai and Ng (JoE, 2013).
- Independence of $\{\mathbf{u}_t\}_{t\geq 1}$ and $\{e_t\}_{t\geq 1}$ can be relaxed.
  - We only require $E(\mathbf{u}_t \mid x_t) = 0$.

Assumptions-4: Errors and Dependence

Assumption

There exists a positive constant $\mathcal{M} < \infty$, independent of N and T, such that:

1. $E(\mathbf{u}_t) = \mathbf{0}$, *and* $E|u_{it}|^8 \leq \mathcal{M}$.
2. $\|\Sigma_u\|_1 \leq \mathcal{M}$, *and for every* $i, j, t, s > 0, (NT)^{-1} \sum_{i,j,t,s} |E(u_{it} u_{js})| \leq \mathcal{M}$
3. *For every* $(t, s), E\left|N^{-1/2}(\mathbf{u}_s' \mathbf{u}_t - E(\mathbf{u}_s \mathbf{u}_t))\right|^4 \leq \mathcal{M}$.

- $\|\Sigma_u\|_1$ is maximum absolute column sum.
- This assumption is the same as Assumption 3.3 of Fan et al (JoE, 2017).
- Conditions are the same as Bai(ECTA, 2003).

Assumptions-5: Kernel, Smoothness, Moments, Bandwidth

- **1. Smoothness of $m(\cdot)$:** $m(\boldsymbol{w})$ is twice continuously differentiable, and the second derivatives are bounded:

$$\sup_{\boldsymbol{w}} \left| \frac{\partial^2 m(\boldsymbol{w})}{\partial w_i \partial w_j} \right| < \infty, \quad \text{for all } i, j \in \{1, \ldots, L\}.$$

- **2. Stationarity:** The process $\{(\boldsymbol{w}_t, e_t)\}$ is strictly stationary and ergodic.

- **3. Mixing Condition:** The sequence $\{(\boldsymbol{w}_t, e_t)\}$ satisfies an α-mixing condition with mixing coefficients $\alpha(k)$ that decay sufficiently fast, i.e. for some $\delta > 0$:

$$\sum_{k=1}^{\infty} \alpha(k)^{\delta/(2+\delta)} < \infty$$

Assumptions-5: Kernel, Smoothness, Moments, Bandwidth

- **4. Moment Conditions:** The error term e_t has finite second moment $\mathbb{E}[e_t^2] = \sigma^2$ and may follow an autoregressive process. The covariates $\boldsymbol{w}_t$ have bounded moments of order $2 + \delta'$ for some $\delta' > 0$:

$$\mathbb{E}[\|\boldsymbol{w}_t\|^{2+\delta'}] < \infty.$$

- **5. Kernel Function:** The kernel function $K(\cdot)$ is a symmetric, bounded, and integrable function with compact support, satisfying:

$$\int K(\psi)d\psi = 1, \quad \int \psi K(\psi)d\psi = 0, \quad 0 < \int \psi\psi^\top K(\psi)d\psi = \boldsymbol{\kappa}_2 < \infty.$$

- **6. Bandwidth:** h is chosen such that $Th^{L+4} \to 0$ as $T \to \infty$.

$$h \to 0, \quad Th^L \to \infty \quad \text{as} \quad T \to \infty.$$

Consistency of Intermediate Steps

- **Lemma-1** Define $\omega_{N,T} = N^{-1/2} + T^{-1/2}$. Under Assumptions 2.1,2.2, 3, 4:

$$\frac{1}{T}\sum_{t=1}^{T} ||\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t||^2 = O_p(\omega_{N,T}^2)$$

  - *Proof:* This result is proved in Theorem 1 of Bai and Ng (ECTA, 2002).

Consistency of Intermediate Steps

- **Lemma-1** Define $\omega_{N,T} = N^{-1/2} + T^{-1/2}$. Under Assumptions 2.1,2.2, 3, 4:

$$\frac{1}{T}\sum_{t=1}^{T} ||\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t||^2 = O_p(\omega_{N,T}^2)$$

  - *Proof:* This result is proved in Theorem 1 of Bai and Ng (ECTA, 2002).

- **Lemma-2** Under Assumptions 2.1,2.2,2.3, 3, and 4:

$$\left\|\widehat{\Sigma}_{\mathbf{f}|x} - \Sigma_{\mathbf{f}|x}\right\| = O_p\left(\omega_{N,T}\right) \qquad \text{and} \qquad \left\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\right\| = O_p\left(\omega_{N,T}\right)$$

  for $j = 1, \ldots, L$, where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L$ form an orthonormal basis for $S_{x|\mathbf{f}}$.
  - *Proof:* Theorem 3.1 of Fan et al (JoE, 2017). Replace their x_{it} with our z_{it} and their y_{t+1} with our x_t.

- *Corollary-1* Under the same conditions of Lemma-2, for any $j = 1, 2, \ldots L$:

$$\widehat{\boldsymbol{\theta}}_j'\widehat{\mathbf{f}}_t \xrightarrow{p} \boldsymbol{\theta}_j'\mathbf{f}_t$$

Consistency of Intermediate Steps

- **Lemma-3:** Under the Assumptions-5:
  - **Bias**:
$$\mathbb{E}[\widehat{m}(\boldsymbol{w})] - m(\boldsymbol{w}) = \frac{1}{2}h^2 \mathrm{tr}\left(\boldsymbol{G}\right) + o(h^2),$$
    where $\boldsymbol{G} = \nabla^2 m(\boldsymbol{w}) \cdot \int \boldsymbol{\psi}\boldsymbol{\psi}^\top K(\boldsymbol{\psi})d\boldsymbol{\psi}$, $\nabla^2 m(\boldsymbol{w})$ is the Hessian matrix of second derivatives of $m(\boldsymbol{w})$.
  - **Variance**:
$$\mathrm{Var}(\widehat{m}(\boldsymbol{w})) = \frac{\sigma^2}{Th^L g(\boldsymbol{w})} \int K^2(\boldsymbol{\psi})d\boldsymbol{\psi} + o\left(\frac{1}{Th^L}\right),$$
    where $g(\boldsymbol{w})$ is the joint density of the covariates $\boldsymbol{w}_t$ at point $\boldsymbol{w}$, and σ^2 is the variance of the error term e_t.

Proof: See Masry (SPA, 1996) and Fan and Gijbels(Routledge, 2018).

Consistency of β

- **Theorem-1:** Define $\delta_{NT} = \min\{N^{1/2}, T^{2/(L+4)}\}$. Under Assumptions 1-5,

$$\left(\widehat{\beta} - \beta\right) = O_p(\delta_{NT}^{-1})$$

Proof: Online Link to Proof

Consistency of β

- **Theorem-1:** Define $\delta_{NT} = \min\{N^{1/2}, T^{2/(L+4)}\}$. Under Assumptions 1-5,

$$\left(\widehat{\beta} - \beta\right) = O_p(\delta_{NT}^{-1})$$

Proof: Online Link to Proof

- Define:

$$\Delta = \begin{bmatrix} 0 & \sum_{t=1}^{T}(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t)) \\ \sum_{t=1}^{T}(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t)) & \sum_{t=1}^{T}(\widehat{m}(\widehat{\boldsymbol{w}}_t)^2 - m(\boldsymbol{w}_t)^2) \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0 \\ \sum_{t=1}^{T}(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t))y_t \end{bmatrix} \quad \text{and} \quad Q_t = [1 \quad m(\boldsymbol{w}_t)]$$

- Then,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = & \left(Q'Q\right)^{-1} Q'e\beta_1 + \left(Q'Q\right)^{-1} Q'\varepsilon - \left(Q'Q\right)^{-1} \Delta \left(Q'Q + \Delta\right)^{-1} \Gamma \\ & + \left(Q'Q\right)^{-1} \Gamma - \left(Q'Q\right)^{-1} \Delta \left(Q'Q + \Delta\right)^{-1} Q'Q\boldsymbol{\beta} \\ & - \left(Q'Q\right)^{-1} \Delta \left(Q'Q + \Delta\right)^{-1} Q'e\beta_1 - \left(Q'Q\right)^{-1} \Delta \left(Q'Q + \Delta\right)^{-1} Q'\varepsilon \\ = & O_p(T^{\frac{-1}{2}}) + O_p(T^{\frac{-1}{2}}) + O_p(\delta_{NT}^{-2}) + O_p(\delta_{NT}^{-1}) \\ & + O_p(\delta_{NT}^{-1}) + O_p(\delta_{NT}^{-1})O(T^{-1/2}) + O_p(\delta_{NT}^{-1})O(T^{-1/2})\end{aligned}$$

Asymptotic Normality of β

- **Theorem-2:** Denote $\mathscr{B} = (Q'Q)^{-1}\Gamma - (Q'Q)^{-1}\Delta(Q'Q+\Delta)^{-1}Q'Q\beta$ where $Q_t = [1 \quad m(\boldsymbol{w}_t)]$. Then, under Assumption-1-5, we have,

$$\delta_{NT}(\widehat{\beta} - \beta - \mathscr{B}) \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{1}{T}Q'Q\right)^{-1} \frac{\mathbf{V}}{T^2\delta_{NT}^{-2}} \left(\frac{1}{T}Q'Q\right)^{-1}\right),$$

where,

  - $\mathbf{V}_{11} = \beta_1^2\left(\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t))\right)^2$
  - $\mathbf{V}_{12} = \mathbf{V}_{21} = \beta_1\left(\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t))\right)\Big[\beta_0\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t))$
$$+\beta_1\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t)^2 - m(\boldsymbol{w}_t)^2)\Big]$$

- $\mathbf{V}_{22} =$
$$\left[\beta_0\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t)) + \beta_1\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t)^2 - m(\boldsymbol{w}_t)^2)\right]^2 + \left(\sum_{t=1}^T(\widehat{m}(\widehat{\boldsymbol{w}}_t) - m(\boldsymbol{w}_t))y_t\right)^2$$

Asymptotic Normality of β: Proof

- From proof of consistency:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(Q'Q\right)^{-1}\Gamma - \left(Q'Q\right)^{-1}\Delta\left(Q'Q+\Delta\right)^{-1}Q'Q\boldsymbol{\beta} + o_p(\delta_{NT}^{-1})$$

-

$$\delta_{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{T}Q'Q\right)^{-1}\frac{1}{T\delta_{NT}^{-1}}\left[\Gamma - \Delta\left(Q'Q+\Delta\right)^{-1}Q'Q\boldsymbol{\beta}\right] + o_p(1)$$

-

$$\text{Var}\left(\delta_{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right) = \left(\frac{1}{T}Q'Q\right)^{-1}\frac{1}{T^2\delta_{NT}^{-2}}\left(\Gamma\Gamma' + \Delta\boldsymbol{\beta}\boldsymbol{\beta}'\Delta'\right)\left(\frac{1}{T}Q'Q\right)^{-1}$$

[Online Link to Proof]

Simulation Design-I: Gains from Supervision

- Design-I:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t && (10)\\
x_t &= \phi' \mathbf{f}_t + e_t && (11)\\
z_{it} &= \mathbf{b}_i' \mathbf{f}_t + \sigma_z u_{it}, \quad 1 \leq i \leq N, 1 \leq t \leq T && (12)\\
f_{jt} &= \gamma_j f_{jt-1} + v_{jt}, \quad 1 \leq j \leq r && (13)
\end{aligned}
$$

- $\beta_0 = 0$, $\beta_1 = 2$, $r = 5$ and $\phi = (0.8, 0.5, 0.3, 0, 0)'$.
- (ε_t, e_t) are generated from the following processes:

$$
\begin{aligned}
\varepsilon_t &= \alpha_1 \varepsilon_{t-1} + \eta_t \\
e_t &= \alpha_2 e_{t-1} + \zeta_t
\end{aligned}
$$

- We control the endogeneity by a parameter $\rho = cor(\eta_t, \zeta_t)$.
- $\gamma_j = \alpha_1 = \alpha_2 = 0.5$ for all j. Loadings $\mathbf{b}_i \sim U[1, 2]$.
- u_{jt} and v_{jt} are generated from $\mathcal{N}(0, 1)$.

Design-I: Results

ρ	r	T	N	$E(\widehat{\beta_1})$ SIF	OLS	2SLS	FIV	RMSE$(\widehat{\beta}_1)$ SIF	OLS	2SLS	FIV
0	5	100	25	2.06	2.06	2.01	2.01	0.38	0.15	0.27	0.28
0	5	100	75	2.07	2.06	2.06	2.06	0.36	0.15	0.15	0.15
0	5	100	125	2.07	2.06	-	1.88	0.37	0.15	-	0.59
0	5	200	50	2.05	2.06	2.05	2.04	0.27	0.11	0.11	0.11
0	5	200	150	2.05	2.06	2.06	2.06	0.27	0.11	0.11	0.11
0	5	200	250	2.05	2.06	-	1.96	0.27	0.11	-	0.38
0	5	400	100	2.03	2.05	2.04	2.05	0.18	0.09	0.09	0.09
0	5	400	300	2.02	2.05	2.05	2.05	0.18	0.09	0.09	0.10
0	5	400	500	2.02	2.05	-	1.98	0.18	0.09	-	0.28
0.5	5	100	25	2.13	2.37	2.23	2.26	0.39	0.41	0.38	0.41
0.5	5	100	75	2.13	2.37	2.50	2.52	0.39	0.41	0.53	0.55
0.5	5	100	125	2.12	2.37	-	1.80	0.38	0.41	-	0.65
0.5	5	200	50	2.09	2.33	2.86	2.90	0.29	0.35	0.87	0.92
0.5	5	200	150	2.08	2.33	2.50	2.51	0.28	0.35	0.51	0.52
0.5	5	200	250	2.07	2.33	-	1.93	0.28	0.35	-	0.37
0.5	5	400	100	2.05	2.31	2.88	2.91	0.20	0.32	0.89	0.93
0.5	5	400	300	2.04	2.31	2.45	2.46	0.20	0.32	0.46	0.47
0.5	5	400	500	2.04	2.31	-	1.97	0.20	0.32	-	0.28
0.9	5	100	25	2.20	2.64	2.37	2.42	0.44	0.66	0.47	0.52
0.9	5	100	75	2.22	2.64	2.74	2.76	0.43	0.66	0.76	0.78
0.9	5	100	125	2.22	2.64	-	1.77	0.43	0.66	-	0.74
0.9	5	200	50	2.15	2.61	3.01	3.01	0.31	0.62	1.02	1.06
0.9	5	200	150	2.12	2.61	2.74	2.75	0.30	0.62	0.75	0.76
0.9	5	200	250	2.12	2.61	-	1.92	0.29	0.62	-	0.38
0.9	5	400	100	2.07	2.59	3.02	3.04	0.21	0.59	1.03	1.06
0.9	5	400	300	2.07	2.59	2.71	2.71	0.20	0.59	0.71	0.72
0.9	5	400	500	2.06	2.59	-	1.96	0.20	0.59	-	0.28

Design-I: Takeaways

- SIF: Least RMSE and less bias in the majority of the cases.
- Bias and RMSE approaches zero as the sample size T.
- No endogeneity ($\rho = 0$) $\implies$ OLS is the best method (in RMSE sense).
- For low N and $\rho \neq 0$: 2SLS does better than OLS.
- Increase in N relative to T hardly affects our estimate, while it appears to do so to other competing methods.
- Increase in sample-size T reduces the bias of all the methods.

Design-II: Gains from Dimension Reduction

- Design-II:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{14}$$
$$x_t = \phi' \mathbf{f}_t + e_t \tag{15}$$
$$z_{it} = \mathbf{b}_i' \mathbf{f}_t + \sigma_z u_{it}, \quad 1 \leq i \leq N, 1 \leq t \leq T \tag{16}$$
$$f_{jt} = \gamma_j f_{jt-1} + v_{jt}, \quad 1 \leq j \leq r \tag{17}$$

- $\beta_0 = 0$, $\beta_1 = 2$, $r = 3$ and $\phi = (0.8, 0.5, 0.3)'$.
- (ε_t, e_t) are generated from the following processes:

$$\varepsilon_t = \alpha_1 \varepsilon_{t-1} + \eta_t$$
$$e_t = \alpha_2 e_{t-1} + \zeta_t$$

- We control the endogeneity by a parameter $\rho = cor(\eta_t, \zeta_t)$.
- $\gamma_j = \alpha_1 = \alpha_2 = 0.5$ for all j. Loadings $\mathbf{b}_i \sim U[1, 2]$.
- u_{jt} and v_{jt} are generated from $\mathcal{N}(0, 1)$.

Design-II: Results

ρ	r	T	N	$E(\widehat{\beta_1})$ SIF	OLS	2SLS	FIV	RMSE$(\widehat{\beta}_1)$ SIF	OLS	2SLS	FIV
0	3	100	25	2.06	2.06	2.00	2.01	0.37	0.15	0.28	0.29
0	3	100	75	2.07	2.06	2.06	2.06	0.37	0.15	0.15	0.15
0	3	100	125	2.07	2.06	-	1.93	0.37	0.15	-	0.42
0	3	200	50	2.04	2.06	2.05	2.04	0.27	0.11	0.11	0.11
0	3	200	150	2.04	2.06	2.06	2.06	0.27	0.11	0.11	0.11
0	3	200	250	2.04	2.06	-	1.97	0.27	0.11	-	0.29
0	3	400	100	2.02	2.05	2.04	2.05	0.18	0.09	0.09	0.09
0	3	400	300	2.02	2.05	2.05	2.05	0.18	0.09	0.09	0.10
0	3	400	500	2.02	2.05	-	1.99	0.18	0.09	-	0.20
0.5	3	100	25	2.10	2.37	2.22	2.28	0.38	0.41	0.37	0.43
0.5	3	100	75	2.10	2.37	2.51	2.52	0.38	0.41	0.54	0.56
0.5	3	100	125	2.10	2.37	-	1.89	0.37	0.41	-	0.41
0.5	3	200	50	2.06	2.33	2.88	2.90	0.27	0.35	0.89	0.95
0.5	3	200	150	2.06	2.33	2.51	2.51	0.27	0.35	0.52	0.53
0.5	3	200	250	2.06	2.33	-	1.96	0.27	0.35	-	0.28
0.5	3	400	100	2.03	2.31	2.89	2.92	0.19	0.32	0.90	0.94
0.5	3	400	300	2.03	2.31	2.45	2.46	0.19	0.32	0.46	0.47
0.5	3	400	500	2.03	2.31	-	1.98	0.19	0.32	-	0.21
0.9	3	100	25	2.15	2.64	2.36	2.42	0.41	0.66	0.46	0.54
0.9	3	100	75	2.16	2.64	2.75	2.77	0.41	0.66	0.77	0.79
0.9	3	100	125	2.17	2.64	-	1.88	0.41	0.66	-	0.42
0.9	3	200	50	2.09	2.61	3.02	3.01	0.28	0.62	1.03	1.08
0.9	3	200	150	2.09	2.61	2.75	2.75	0.28	0.62	0.76	0.76
0.9	3	200	250	2.09	2.61	-	1.95	0.28	0.62	-	0.28
0.9	3	400	100	2.05	2.59	3.03	3.02	0.20	0.59	1.04	1.07
0.9	3	400	300	2.05	2.59	2.71	2.71	0.21	0.59	0.71	0.72
0.9	3	400	500	2.05	2.59	-	1.97	0.20	0.59	-	0.21

Design-III: Gains from Handling Non-linearities

- Design-III:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t && (18) \\
x_t &= f_{1t}\left(f_{2t} + f_{3t} + 1\right) + e_t && (19) \\
z_{it} &= \mathbf{b}_i'\mathbf{f}_t + \sigma_z u_{it}, \quad 1 \le i \le N, 1 \le t \le T && (20) \\
f_{jt} &= \gamma_j f_{jt-1} + v_{jt}, \quad 1 \le j \le r && (21)
\end{aligned}
$$

- $\beta_0 = 0$, $\beta_1 = 2$, $r = 3$ and $\phi_1 = (1, 0, 0)'$, $\phi_2 = (0, 1/\sqrt{2}, 1/\sqrt{2})'$.
- (ε_t, e_t) are generated from the following processes:

$$
\begin{aligned}
\varepsilon_t &= \alpha_1 \varepsilon_{t-1} + \eta_t \\
e_t &= \alpha_2 e_{t-1} + \zeta_t
\end{aligned}
$$

- We control the endogeneity by a parameter $\rho = cor(\eta_t, \zeta_t)$.
- $\gamma_j = \alpha_1 = \alpha_2 = 0.5$ for all j. Loadings $\mathbf{b}_i \sim U[1, 2]$.
- u_{jt} and v_{jt} are generated from $\mathcal{N}(0, 1)$.

Design-III: Results

ρ	r	T	N	$E(\widehat{\beta_1})$ SIF	OLS	2SLS	FIV	RMSE$(\widehat{\beta}_1)$ SIF	OLS	2SLS	FIV
0	3	100	25	2.24	2.05	2.01	2.02	0.46	0.16	0.30	0.32
0	3	100	75	2.23	2.05	2.05	2.05	0.43	0.16	0.18	0.18
0	3	100	125	2.23	2.05	-	1.67	0.42	0.16	-	2.02
0	3	200	50	2.20	2.05	2.06	2.05	0.31	0.12	0.13	0.14
0	3	200	150	2.19	2.05	2.05	2.05	0.30	0.12	0.13	0.13
0	3	200	250	2.19	2.05	-	1.87	0.30	0.12	-	0.91
0	3	400	100	2.14	2.04	2.05	2.05	0.20	0.09	0.11	0.11
0	3	400	300	2.14	2.04	2.04	2.04	0.20	0.09	0.10	0.10
0	3	400	500	2.14	2.04	-	1.93	0.20	0.09	-	0.53
0.5	3	100	25	2.28	2.28	2.23	2.26	0.46	0.33	0.41	0.44
0.5	3	100	75	2.28	2.28	2.41	2.42	0.46	0.33	0.46	0.48
0.5	3	100	125	2.28	2.28	-	1.57	0.46	0.33	-	1.83
0.5	3	200	50	2.22	2.25	2.93	2.96	0.33	0.28	0.95	1.00
0.5	3	200	150	2.22	2.25	2.41	2.41	0.32	0.28	0.43	0.43
0.5	3	200	250	2.21	2.25	-	1.87	0.32	0.28	-	1.05
0.5	3	400	100	2.15	2.23	2.94	2.97	0.22	0.24	0.95	0.99
0.5	3	400	300	2.15	2.23	2.36	2.37	0.22	0.24	0.38	0.38
0.5	3	400	500	2.15	2.23	-	1.92	0.21	0.24	-	0.55
0.9	3	100	25	2.35	2.47	2.36	2.41	0.54	0.51	0.49	0.53
0.9	3	100	75	2.35	2.47	2.60	2.62	0.50	0.51	0.64	0.66
0.9	3	100	125	2.36	2.47	-	1.59	0.51	0.51	-	1.98
0.9	3	200	50	2.25	2.45	3.10	3.10	0.34	0.46	1.12	1.16
0.9	3	200	150	2.25	2.45	2.60	2.61	0.34	0.46	0.62	0.62
0.9	3	200	250	2.25	2.45	-	1.84	0.34	0.46	-	0.92
0.9	3	400	100	2.17	2.43	3.11	3.12	0.23	0.44	1.12	1.16
0.9	3	400	300	2.17	2.43	2.57	2.57	0.23	0.44	0.57	0.58
0.9	3	400	500	2.17	2.43	-	1.90	0.22	0.44	-	0.56

Design-III: Takeaways

- Qualitatively the results are the same as previous designs.
- Gains of capturing non-linearity are not that visible for smaller sample sizes.

Design-4: Middle Ground for Sparse and Dense

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (22)$$

$$x_t = \phi' \mathbf{f}_t + e_t \quad (23)$$

$$z_{it} = \mathbf{b}_i' \mathbf{f}_t + \sigma_z u_{it} \qquad 1 \leq i \leq N/2, \qquad 1 \leq t \leq T \quad (24)$$

$$z_{it} = \sigma_z (e_t + u_{it}) \qquad N/2 + 1 \leq i \leq 5N/8, \qquad 1 \leq t \leq T \quad (25)$$

$$z_{it} = \sigma_z u_{it} \qquad 5N/8 + 1 \leq i \leq N, \qquad 1 \leq t \leq T \quad (26)$$

$$f_{jt} = \gamma_j f_{jt-1} + v_{jt} \qquad 1 \leq j \leq r, \qquad 1 \leq t \leq T \quad (27)$$

- $\beta_0 = 0$, $\beta_1 = 2$, $r = 3$ and $\phi = (0.8, 0.5, 0.3)'$.
- Eq-24 generates N/2 dense instruments.
- Eq-25 and Eq-26 generates N/2 sparse instruments.

Design-4: Results

			$E(\widehat{\beta}_1)$			$RMSE(\widehat{\beta}_1)$		
ρ	T	N	SIF	FIV	PLIV	SIF	FIV	PLIV
0	200	50	2.10	2.06	2.08	0.12	0.11	0.11
0	200	150	2.10	2.06	2.09	0.13	0.11	0.12
0	200	250	2.11	2.06	2.09	0.13	0.11	0.12
0.5	200	50	2.52	2.46	2.50	0.54	0.48	0.51
0.5	200	150	2.52	2.37	2.46	0.53	0.38	0.47
0.5	200	250	2.52	2.33	2.45	0.53	0.35	0.46
0.9	200	50	2.91	2.83	2.89	0.91	0.84	0.90
0.9	200	150	2.92	2.67	2.84	0.93	0.68	0.84
0.9	200	250	2.92	2.61	2.83	0.93	0.62	0.83

Empirical Exercise I: BLP

- **Question**: Estimation of price elasticity of automobile demand.
  - When we use OLS, we get elasticity to be either zero or positive.
  - *BLP (ECTA, 1995)*-equation:

$$y_{it} = \alpha_0 p_{it} + x_{it}' \beta_0 + \varepsilon_{it} \tag{28}$$
$$p_{it} = z_{it}' \delta_0 + x_{it}' \gamma_0 + u_{it} \tag{29}$$

    - BLP-paper found elasticity to be -0.14.
  - Assuming sparsity Chernozhukov et al. (AER, 2015) finds elasticity$=-0.18$.
- We assume instruments are dense.

$$z_{it} = \lambda_i f_t + \nu_{it}$$

  - Are they dense?
    - Using Chernozhukov et al. (AER, 2015), we show that two factors can explain 85% of the variation in the instrument set.

BLP: Price Elasticity of Automobile Demand

- Instrument Validity Check:
  - *Relevancy:*
    - Our true instruments are $\boldsymbol{\theta}_1'\mathbf{f}_t, \dots, \boldsymbol{\theta}_L'\mathbf{f}_t$ are well correlated with price.
  - *Exclusion:*
    - Berry et al (ECTA, 1995) provides arguments for z satisfying exclusion.
    - We use common components of z. If z are valid instruments, a portion of z should also be valid.
- Data:
  - We use the same data used in Chernozhukov et al. (AER, 2015).
- Results:
  - We find price elasticity= -0.152.   95% C.I.=[-0.165, -0.125].
  - Our results confirm previous studies.

Empirical Exercise II: CAPM-Beta

- y_t: Returns on BlackRock's SmallCap ETF (IJS). x_t: Market returns.
- Where is endogeneity? :
  - R_t^* is the return on the market portfolio (not observed).

$$y_t = \alpha + \beta R_t^* + \eta_t \tag{30}$$

  - Since R^* is not observed, we use its proxy $R = R^* + e_t$:

$$y_t = \alpha + \beta R_t + \varepsilon_t$$

  - $\varepsilon_t = e_t - \beta\eta_t$ is correlated with R_t: endogeneity.
- Instruments:
  - Systemic factors driving economic sentiments are our instruments.
    - We use returns on S&P500 firms (expect ones in DJIA) as z_{it}.
  - Return on DJIA is our R_t and return on IJS is our y_t.

CAPM-Beta

- Instrument Validity:
  - *Relevancy:*
    - We extract common factors from the set of z which correlate well with R_t.
  - *Exclusion:*
    - IJS (index of small firms) is unlikely to drive systemic risks in the economy.
- Data:
  - We use CRSP and Wall Street Journal to get daily data.
  - Sample period: 2001 to 2023.
- Results:
  - CAPM Beta computed from our method is 1.54.
    - We reject null that IJS Beta=1.
  - OLS computes IJS Beta to be near 1.

Conclusion

- Proposed a procedure to sufficiently filter information from many instruments.
- Four attractive features of the method:
  - Weaker Relevancy Assumptions.
  - Supervision.
  - Dimension Reduction.
  - Ability to handle non-linearities.
- Simulation exercises confirm the aforementioned claimed properties.
- Two empirical exercises are considered as a proof of concept.