Recent Advances in Nonparametric Instrumental Regression

J.S. Racine

Department of Economics,
Department of Mathematics and Statistics
McMaster University
Hamilton, Ontario, Canada

September 26 2014, AHEW 2014, U of T

Support from SSHRC, NSERC, and Compute Canada is gratefully acknowledged

Overview

- Instrumental regression defined
- Finite- versus infinite-dimensional models
- Ill-posed Inverse Problems
- Identification
- Nonparametric density estimation - an ill-posed inverse problem
- Nonparametric instrumental regression - an ill-posed inverse problem
- Regularization methods
- Illustrations
- Availability of implementations

What are Instrumental Regression Methods?

- 'Instrumental regression' methods are econometric techniques for estimating causal relationships on the basis of observational (i.e. non-experimental) data
- The problem arises when one (or more) of the explanatory variables is known to be correlated with the stochastic error process in the underlying data generating process (DGP)
- When this is the case, standard regression methods (i.e. ones that ignore this correlation) produce biased and inconsistent estimates

Historical Perspective

- If 'valid' instruments exist (i.e. variables that are correlated with the 'endogenous' explanatory variables but are uncorrelated with the error term), then instrumental regression methods may be used
- The earliest solution to this problem in Econometrics is Wright (1928) (cf Stock & Trebbi (2003))
- Wright (1928, Appendix B) showed that instrumental regression can be used to estimate the coefficient on an endogenous regressor, something that e.g. ordinary least squares regression cannot do
- Instrumental regression has emerged as a central technique of modern micro- and macroeconometrics

Finite-Dimensional Simplifications

- Applied researchers often presume that the functional forms of the instrumental regression model (either single- or multi-equation) are known (e.g. linear and additive)
- 'Parametric' estimation methods treat the model as fixed and exact (i.e. 'finite-dimensional')
- Nonparametric estimation methods treat the model as an approximation that depends on the sample size (i.e. 'infinite-dimensional', cf Horowitz (2014))
- Many practical problems require 'nonparametric' estimates

Parametric Versus Nonparametric Instrumental Methods

- Except in special cases, parametric and nonparametric methods may deliver different estimates, confidence intervals, and outcomes of hypothesis tests
- Because parametric estimation assumes a fixed model that does not depend on the sample size, parametric methods typically indicate that the estimates are more precise than they really are
- Consequently, conclusions that are supported by a parametric estimator may not be supported by a nonparametric estimator

Identification

- Let φ denote some 'parameter' of interest, which could be a scalar, vector, or function (i.e. $\varphi = \varphi(Z)$)
- This parameter is 'identified' if it is uniquely determined by the probability distribution from which the available data are sampled
- A trivial illustration would be, say, $\varphi = \int_{-\infty}^{\infty} y\, dF(y)$, the population mean (i.e. provided that the population mean exists, it is identified and can be consistently estimated by its sample counterpart)
- Strictly speaking, a parameter is identified (overidentified) if there is a one-to-one (or many-to-one) mapping from the population distribution to the parameter

Continuous Identifying Mappings

- For many popular parametric estimators (e.g. linear least squares or linear instrumental regression) the parameter of interest is a vector and the 'identifying mapping' is 'continuous'
- By 'continuous' we mean that small changes in the population distribution produce only small changes in the identified parameter
- In such cases, the parameter can be consistently estimated by replacing the unknown population distribution with a consistent sample analog (e.g. the empirical distribution of the data)

Discontinuous Identifying Mappings

- This approach (i.e. replacing the unknown population distribution with a consistent sample analog) can fail if the mapping that identifies the parameter of interest is discontinuous
- When the identifying mapping is discontinuous, the parameter of interest φ cannot be estimated consistently by simply replacing the population distribution of the data with a consistent sample analog
- This occurs because the estimated and true values of φ can be very different even if the sample size is large enough to make the difference between the sample and population distribution negligibly small

Ill-Posed Problems

- Hadamard (1923) called a problem 'well-posed' if it has a unique solution that depends continuously on the available data
- We say that an estimation problem is 'ill-posed' if the identifying mapping is discontinuous in a way that prevents consistent estimation of the parameter of interest by replacing the population distribution of the data with a consistent sample analog
- We say that an estimation problem is an 'ill-posed inverse problem' if the discontinuous identifying mapping is obtained by inverting another mapping that is continuous

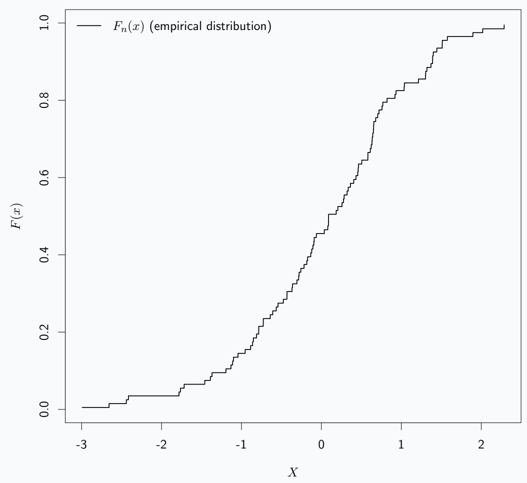
Nonparametric Density Estimation

- Nonparametric estimation of a probability density function ('PDF') presents an ill-posed inverse problem and will illustrate the problem nicely
- Let $F(x) \equiv \int_{-\infty}^{x} f(t)\,dt$ denote a cumulative distribution function ('CDF') and $f(x) \equiv dF(x)/dx$ denote the PDF
- The nonparametric 'empirical' CDF is given by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x),$$

  where $\mathbf{1}(A)$ is an 'indicator function' taking values 1 if A is true, zero otherwise
- $F_n(x)$ is a uniformly consistent estimator of $F(x)$

Example (The Empirical Distribution Function)



Nonparametric Density Estimation

- Although $F_n(x)$ is uniformly consistent, it is a step function with derivative taking values either 0 or ∞ (when $0 < f(x) < \infty$ i.e. is bounded) even as $n \to \infty$
- Nonparametric estimation of $f(x)$ is thus an ill-posed inverse problem
- That is, $f(x)$ cannot be consistently estimated simply by replacing $F(x)$ with a consistent estimator such as $F_n(x)$ in $f(x) = dF(x)/dx$
- For density estimation we adopt a 'regularization' approach (i.e. we modify or "regularize" the identifying mapping) by 'smoothing' $F_n(x)$, with 'regularization parameter' the 'bandwidth' h, as demonstrated below

Nonparametric Density Estimation

- Let the kernel smoothed CDF estimator be

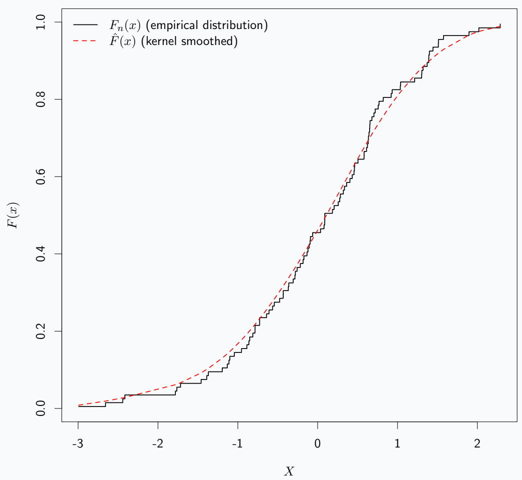
$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{x - X_i}{h}\right)$$

  where Φ is e.g. the Gaussian CDF kernel, h a 'bandwidth'
- The kernel smoothed PDF estimator is

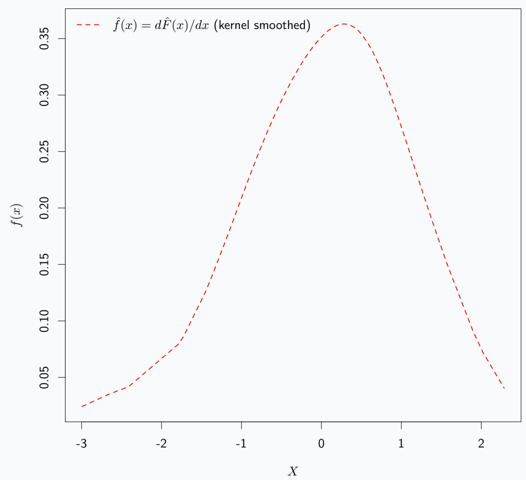
$$\hat{f}(x) = \frac{d\hat{F}(x)}{dx} = \frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x - X_i}{h}\right)$$

  where $\phi = \Phi'$, e.g. the Gaussian PDF kernel
- For consistent estimation of $f(x)$ the bandwidth h must vanish as $n \to \infty$ but not too quickly ($nh \to \infty$ as $n \to \infty$), otherwise there is not enough regularization to overcome the discontinuity

Example (Ill-Posed Inverse Problem - Estimating $f(x) = dF(x)/dx$)



Example (Ill-Posed Inverse Problem - Estimating $f(x) = dF(x)/dx$)



Nonparametric Density Estimation

- To summarize and cement ideas, we have the relationship

$$F(x) = \int_{-\infty}^{\infty} f(t)\mathbf{1}(t \leq x)\,dt$$

- The mapping from $f(x)$ to $F(x)$ is a continuous one (i.e. small changes in $f(x)$ produce small changes in $F(x)$)
- However, the converse is not true (i.e. the inverse mapping from $F(x)$ to $f(x)$ given by $dF(x)/dx$ is not a continuous mapping)
- Therefore replacing $F(x)$ with a consistent estimator such as $F_n(x)$ and attempting to solve for $f(x)$ can fail

Nonparametric Density Estimation

- The fact that the nonparametric model is infinite-dimensional lies at the root of the problem
- This is why nonparametric density estimation constitutes an ill-posed inverse problem and requires a regularized solution
- Kernel smoothing methods are one such approach to overcoming this problem for nonparametric density estimation (not the only one of course)
- What does this have to do with instrumental regression you may ask?
- Quite a lot as it turns out (kernel smoothing methods will be used to consistently estimate certain objects in the 'regularized' solution to avoid committing a 'regularization crime')

Nonparametric Instrumental Regression

- Consider an instrumental regression model given by

$$Y = \varphi(Z) + U$$

where the parameter $\varphi(Z)$ is unknown with $Y, X \in \mathbb{R}$, $E(Y^2) < \infty$
- Assume that $\varphi(Z)$ and its derivative $\varphi'(Z)$ are smooth and square integrable w.r.t. the Lebesgue measure
- The object of interest, $\varphi(z)$, is no longer given by the conditional mean of Y since $E(U|z) \neq 0$, i.e.

$$E(Y|z) = \varphi(z) + E(U|z) \neq \varphi(z)$$

- Given a valid instrument W for which $E(U|w) = 0$, Darolles, Fan, Florens & Renault (2011) recover $\varphi(Z)$ as the solution to

$$E(Y|w) = E(\varphi(Z)|w)$$

Nonparametric Instrumental Regression

- $\varphi(z)$ corresponds to any solution of the integral equation

$$E(Y|w) = \int_{-\infty}^{\infty} \varphi(z) f(z|w)\,dz$$

- The mapping from $\varphi(z)$ to $E(Y|w)$ is continuous provided $f(z|w)$ is bounded
- However, the inverse mapping from $E(Y|w)$ to $\varphi(z)$ is discontinuous, hence estimation of $\varphi(z)$ constitutes an ill-posed inverse problem
- Note the similarity between the above and nonparametric density estimation, another ill-posed inverse problem,

$$F(x) = \int_{-\infty}^{\infty} f(t)\mathbf{1}(t \leq x)\,dt$$

Regularization and Instrumental Regression

- Regularization methods are techniques for removing the discontinuity in the identifying mapping in ill-posed problems in order to facilitate estimation (cf Horowitz (2014))
- For the ill-posed nonparametric density estimation problem we used smoothing as a regularization device
- For nonparametric instrumental regression we use different regularization methods (e.g.Tikhonov (1943) or Landweber (1951)-Fridman (1956))
- Alternative semi- and nonparametric approaches include use of control functions (Newey, Powell & Vella (1999)) and sieve approximations (Newey & Powell (2003), Horowitz (2011), Chen & Pouzo (2012))

Regularization Instrumental Regression

- A regularized solution of the ill-posed instrumental regression problem is a sequence of solutions of a sequence of well-posed problems (called the 'regularized problems') such that the sequence converges to the desired object given minimal assumptions about $\varphi(Z)$
- To avoid overly cumbersome notation, define

$$\begin{aligned} r &\equiv E(Y|w), \\ T\varphi &\equiv E(\varphi(Z)|w) \end{aligned}$$

- T is the conditional expectations operator, so with this notation the instrumental relationship $E(Y|w) = E(\varphi(Z)|w)$ can be concisely written as $r = T\varphi$

Landweber-Fridman Regularization

- Given the identifying relationship $r = T\varphi$, we first take the scalar product with respect to the adjoint operator $T^\dagger$ and constant c ($T^\dagger = E(\cdot|z)$ is the adjoint of $T = E(\cdot|w)$, $c < 1$)) then subtract this from φ and manipulate, i.e.

$$\begin{aligned} r &= T\varphi, && (E(Y|w) = E(\varphi(Z)|w)) \\ cT^\dagger r &= cT^\dagger T\varphi, && \text{(premultiply above by } cT^\dagger) \\ \varphi - cT^\dagger r &= \varphi - cT^\dagger T\varphi, && \text{(subtract above from } \varphi) \\ \varphi &= \varphi - cT^\dagger T\varphi + cT^\dagger r && \text{(add } cT^\dagger r\text{, collect terms)} \\ &= \varphi + cT^\dagger(r - T\varphi) \end{aligned}$$

- We then pursue an iterative scheme to solve this well-posed regularized problem for φ

Solving the Regularization Problem

- The iterative scheme is of the form

$$\varphi_k = \varphi_{k-1} + cT^\dagger(r - T\varphi_{k-1}),$$

which, using the original conditional expectation notation, is

$$\varphi_k(z) = \varphi_{k-1}(z) + cE\left[E(Y - \varphi_{k-1}(Z)|w)|z\right]$$

- To obtained the regularized solution for $\varphi(z)$, we estimate the unknown conditional mean objects nonparametrically and iterate
- In order to begin, we require a starting value for $\varphi(z)$, so we could use e.g. $\hat\varphi_0(z) = \widehat{E}(Y|z)$ (the nonparametric conditional mean)

Solving the Regularization Problem

- Next, given $\hat{\varphi}_0(z)$, we can compute $\widehat{E}(Y - \hat{\varphi}_0(Z)|w)$
- Next, we can compute $\widehat{E}\left[\widehat{E}(Y - \hat{\varphi}_0(Z)|w)|z\right]$
- This gives us $\hat{\varphi}_1(z)$ since $\hat{\varphi}_1(z) = \hat{\varphi}_o(z) + c\widehat{E}\left[\widehat{E}(Y - \hat{\varphi}_0(Z)|w)|z\right]$
- Now we can use $\hat{\varphi}_1(z)$ to compute $\hat{\varphi}_2(z)$ in a similar manner thereby obtaining the sequence of solutions $\hat{\varphi}_0(z)$, $\hat{\varphi}_1(z)$, $\hat{\varphi}_2(z)$,...
- This process continues until some 'stopping rule', e.g. $\|(\widehat{E}(Y|w) - \widehat{E}(\hat{\varphi}_k(Z)|w))/\widehat{E}(Y|w)\|^2$, stabilizes from iteration to iteration (the stopping rule is playing the role of the regularization parameter)

Simulated Illustration

Let's consider drawing a sample of size $n = 1000$ for the DGP used by Darolles et al. (2011) which is given by

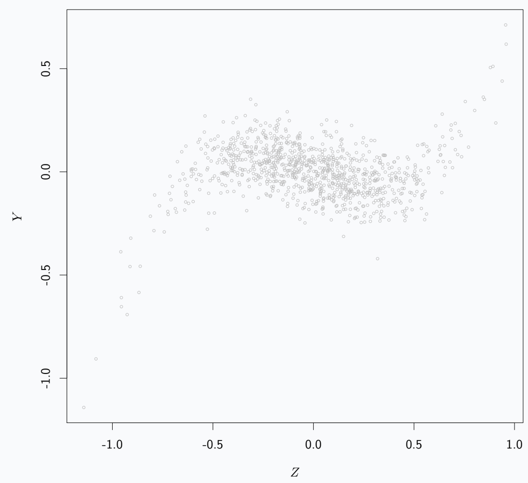
Example (DGP used by Darolles et al (2011))

$$Y = \varphi(Z) + U$$
$$Z = \rho_{z,w} W + V$$
$$U = \rho_{u,z} V + \epsilon$$

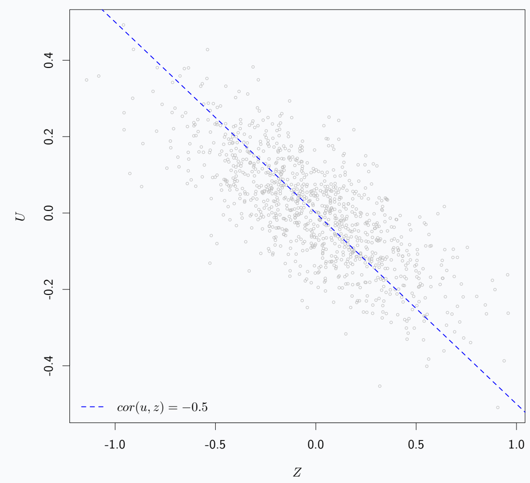
where $\rho_{u,z} = -0.5$, $\rho_{z,w} = 0.2$, $\varphi(Z) = Z^3$, and

$$W \sim N(0, 1)$$
$$V \sim N(0, 0.27^2)$$
$$\epsilon \sim N(0, .05^2)$$

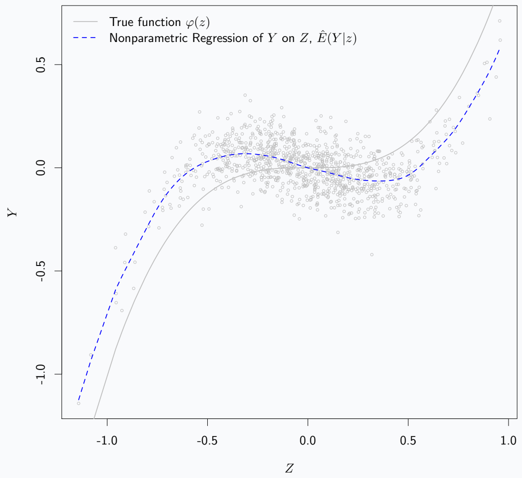
Example (Plot of Y versus (endogenous) Z)



Example (Plot of (unobserved) U versus Z)

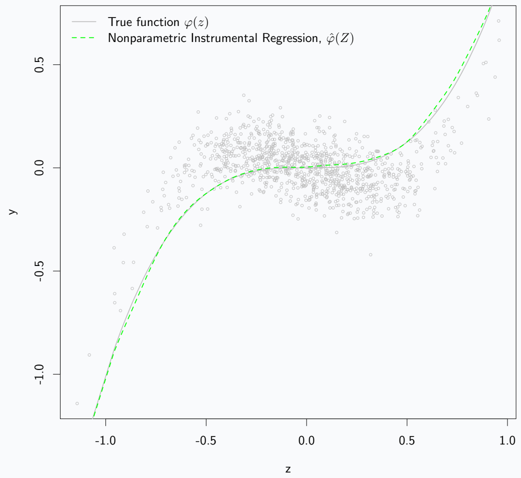


Example (DGP ($\varphi(Z)$) and nonparametric $\hat{E}(Y|Z=z)$)



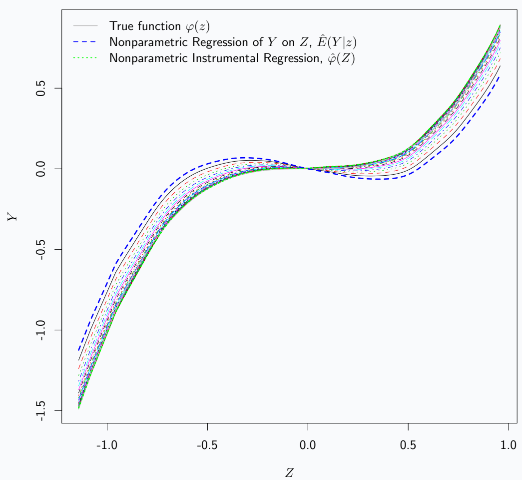
Recent Advances in Nonparametric Instrumental Regression
└Nonparametric IV Regression Illustration

Example (DGP ($\varphi(Z)$) and regularized nonparametric $\hat{\varphi}(Z)$)



Recent Advances in Nonparametric Instrumental Regression
└Nonparametric IV Regression Illustration

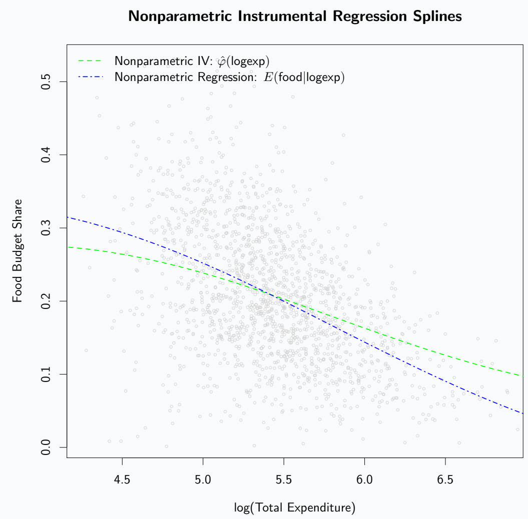
Example (Regularized solution (stopping rule))



Recent Advances in Nonparametric Instrumental Regression
└Nonparametric IV Regression Illustration

- For further details on these methods, see Darolles et al. (2011), Horowitz (2011), and Florens & Racine (2014*a*)
- Consider a random sample taken from the 1995 British Family Expenditure survey used in Blundell, Chen & Kristensen (2007) to estimate an Engel curve
- If Y is a household's expenditure share on a good or service and Z is the household's total expenditure, then $\varphi(Z)$ is an Engel curve (Z and Y are jointly determined)
- If income from wages and salaries is not influenced by household budgeting decisions, then the household head's total earnings from wages and salaries can be used as an instrument, W, for Z (Blundell et al. (2007))

Example (Engel Curve for British Family Expenditure Data)



Nonparametric Models and Instrumental Regression

- Discontinuous identifying relations often arise when the parameter of interest is a function rather than a finite-dimensional quantity (such as a coefficient in a simple parametric specification)
- Infinite-dimensional nonparametric models give rise to ill-posed inverse problems
- Many economists prefer finite-dimensional parametric models for their research
- Economic theory, rarely, if ever provides parametric models
- Economic theory does, however, provide shape constraints (concavity, monotonicity etc.)
- See Du, Parmeter & Racine (2013) and (Florens & Racine (2014*b*)) for work on shape constrained nonparametric (instrumental) regression

Limitations of Instrumental Regression

- Finding a good instrument (like achieving nirvana) is known not to be an easy task (instruments, though correlated with the endogenous regressor, might still be correlated with the error process)
- Even with a valid instrument, the instrument may be 'weak' (the degree of correlation with the endogenous predictor may be low)
- For nonparametric approaches, how to best choose smoothing parameters remains an ongoing active area of research
- We are not yet at the point where these methods are anywhere near as complete as standard nonparametric regression analysis, but these methods will be entering the mainstream in the near future (lots to do but stay tuned!)

Software Implementations

- 'Beta' versions of some functions exist in the `R` packages `np` and `crs` (Hayfield & Racine (2008), Nie & Racine (2012))
- See the functions `npregiv` and `npregivderiv` in the `np` package (nonparametric kernel implementation)
- See the functions `crsiv` and `crsivderiv` in the `crs` package (semi- and nonparametric spline/sieve implementation)
- These functions are quite general, allow for discrete and continuous variables and multiple endogenous, exogenous, and instruments
- The 1995 British Family Expenditure survey data can be found in both packages (`Engel95`)
- The DGP underlying the simulated examples can also be found in both packages (see e.g. `?npregiv` or `?crsiv`)

References I

Blundell, R., Chen, X. & Kristensen, D. (2007), 'Semi-nonparametric IV estimation of shape invariant Engel curves', *Econometrica* **75**, 1613–1669.

Chen, X. & Pouzo, D. (2012), 'Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals', *Econometrica* **80**, 277–321.

Darolles, S., Fan, Y., Florens, J. P. & Renault, E. (2011), 'Nonparametric instrumental regression', *Econometrica* **79**(5), 1541–1565.

Du, P., Parmeter, C. F. & Racine, J. S. (2013), 'Nonparametric kernel regression with multiple predictors and multiple shape constraints', *Statistica Sinica* **23**(3), 1343–1372.

References II

Florens, J. & Racine, J. (2014*a*), Nonparametric instrumental derivatives, Technical report, Université Toulouse 1 Capitole & McMaster University.

Florens, J. & Racine, J. (2014*b*), Shape constrained nonparametric instrumental regression, Technical report, Université Toulouse 1 Capitole & McMaster University.

Fridman, V. M. (1956), 'A method of successive approximations for fredholm integral equations of the first kind', *Uspeskhi, Math. Nauk.* **11**, 233–334.

Hadamard, J. (1923), *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale University Press.

Hayfield, T. & Racine, J. S. (2008), 'Nonparametric econometrics: The np package', *Journal of Statistical Software* **27**(5).
URL: *http://www.jstatsoft.org/v27/i05/*

References III

Horowitz, J. (2011), 'Applied nonparametric instrumental variables estimation', *Econometrica* **79**, 347–394.

Horowitz, J. (2014), 'Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter', *Journal of Econometrics* **180**, 158–173.

Landweber, L. (1951), 'An iterative formula for fredholm integral equations of the first kind', *American Journal of Mathematics* **73**, 615–624.

Newey, W. K. & Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Newey, W. K., Powell, J. L. & Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**, 565–603.

References IV

Nie, Z. & Racine, J. S. (2012), 'The crs package: Nonparametric regression splines for continuous and categorical predictors', *The R Journal* **4**(2), 48–56.

Stock, J. & Trebbi, F. (2003), 'Who invented instrumental variable regression?', *Journal of Economic Perspectives* **17**(3), 177–194.

Tikhonov, A. N. (1943), 'On the stability of inverse problems', *Doklady Akademii Nauk SSSR* **39**(5), 195–198.

Wright, P. G. (1928), *The Tariff on Animal and Vegetable Oils*, Macmillan.

Appendix: Parametric Instrumental Regression

- Consider the finite-dimensional parametric model where we assume that $\varphi(Z) = Z\beta$, hence

$$Y = Z\beta + U,$$

where Z and U are correlated, $E(Y^2) < \infty$ and $E(X^2) < \infty$, with instrument W such that $E(W'U) = 0$, hence

$$E(W'Y) = E(W'Z)\beta$$

- Premultiplying by $E(Z'W)[E(W'W)]^{-1}$ yields

$$E(Z'W)[E(W'W)]^{-1}E(W'Y) = E(Z'W)[E(W'W)]^{-1}E(W'Z)\beta$$

Appendix: Parametric Instrumental Regression

- If the inverse matrices exist then

$$\beta = \left\{E(Z'W)\left[E(W'W)\right]^{-1}E(W'Z)\right\}^{-1} E(Z'W)\left[E(W'W)\right]^{-1}E(W'Y),$$

which identifies β

- Furthermore β is a continuous function of the moments and probability distributions of the r.h.s. random variables

Appendix: Parametric Instrumental Regression

- Given data, replacing unknown expectations with sample averages is equivalent to replacing the unknown distribution of (Y, Z, W) with the empirical distribution of the data, hence

$$\hat{\beta}_{IV} = \left\{n^{-1}\sum_{i=1}^{n} Z_i W_i' \left[n^{-1}\sum_{i=1}^{n} W_i W_i'\right]^{-1} n^{-1}\sum_{i=1}^{n} W_i Z_i'\right\}^{-1} \times n^{-1}\sum_{i=1}^{n} Z_i W_i' \left[n^{-1}\sum_{i=1}^{n} W_i W_i'\right]^{-1} n^{-1}\sum_{i=1}^{n} W_i Y_i$$

- $\hat{\beta}_{IV}$ is a consistent estimator for β in the parametric model (but not for $\varphi(Z)$ *unless* $\varphi(Z)$ is in fact your parametric guess, $Z\beta$, a measure zero event)

Appendix: Parametric Instrumental Regression

- This works simply because using a linear parametric specification for $\varphi(Z)$ renders $E(\varphi(Z)|W) = E(W'Z)\beta$ finite-dimensional and non-singular, so we can solve a set of (finitely many) linear equations
- But if we treat this as an infinite-dimensional problem we are, in effect, trying to solve a set of infinitely-many equations in infinitely-many unknowns, so $T = E(\varphi(Z)|w)$ is, roughly speaking, a "nearly singular infinite-dimensional matrix"
- For this reason, we adopt regularized solutions