Identification and estimation in the functional linear instrumental regression*

Andrii Babii†

Toulouse School of Economics

March 9, 2016

First draft: October 9, 2014

This paper studies a particular type of a linear IV regression model with high-dimensional endogenous component, called the functional linear instrumental regression (FLIR). It is shown that identification in this model can be achieved with a single real-valued instrumental variable under the weak completeness condition. Two estimators based on the Tikhonov and Galerkin regularizations are studied. We obtain the non-asymptotic upper bounds on the mean-integrated squared errors and corresponding convergence rates for both estimators. Estimators are simple to implement and demonstrate good small-sample performance in Monte Carlo experiments.

1 Introduction

Consider the scalar dependent variable Y_i which depends on the random function, $Z_i = (Z_i(t))_{t\in\mathscr{T}}$, where the set $\mathscr{T}$ represents time, age, location, or some other dimension,

$$Y_i = \int_{\mathscr{T}} \beta(t) Z_i(t)\mathrm{d}t + U_i, \quad \mathbb{E}[U_i|Z_i] = 0, \qquad i = 1, \dots, n. \tag{1}$$

The model specifies the linear conditional mean function

$$\mathbb{E}[Y_i|Z_i] = \int_{\mathscr{T}} \beta(t) Z_i(t)\mathrm{d}t,$$

where the slope parameter β is a function, which reflects different strength of the effect of the process $Z_i(t)$ at different $t \in \mathscr{T}$. The exogeneity assumption, $\mathbb{E}[U_i|Z_i] = 0$, however, is not appropriate in many applications. Florens and Van Bellegem (2015) suggest treating endogeneity with the functional instrumental variable, which satisfies the unconditional moment restriction. In this paper, we show that the slope parameter can be identified and consistently estimated with a *single real-valued* instrumental variable under the conditional moment restriction, $\mathbb{E}[U_i|W_i] = 0$. This offers wider range of possible applications, where the only available IV is not functional.

*I'm especially thankful to Jean-Pierre Florens for many helpful discussions and suggestions. I'm thankful to Joel Horowitz and participants of the "Conference on Inverse Problems in Econometrics" at Northwestern University. I'm grateful to Mario Rothfelder for careful reading of the manuscript and many useful suggestions. I would also like to thank to Pascal Lavergne, Bruno Biais, Sophie Moinas and Aleksandra Fridman.

†Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). E-mail: babii.andrii@gmail.com.

Our identification strategy relies on the modified version of the *completeness* property of the distribution of (Z_i, W_i). Even when the slope parameter is identified, its estimation is difficult because the identifying equation is an example of the linear ill-posed integral equation. To overcome ill-posedness we introduce two estimators: based on Tikhonov regularization and orthogonal series decomposition and projection. The appropriate amount of regularization is controlled by tuning parameters and should balance the bias and the variance of corresponding estimators in the optimal manner similarly as in nonparametric estimation problems. We obtain convergence rates of two estimators in the mean-integrated squared error (MISE). The speed of convergence depends not only on smoothness properties of the function β, but also on the degree of ill-posedness. It is verified that speed of convergence of the orthogonal series estimator is optimal from the minimax point for both mildly and severely ill-posed problems. All results are valid for the i.i.d. data as well as for the time series under the appropriate weak dependence conditions.

Linear ill-posed problems with estimated integral operators received lots of attention in econometrics and statistics recently, see Carrasco, Florens, and Renault (2007), Hoffmann and Reiss (2008). Other examples of such models include the nonparametric instrumental variables and the deconvolution. Darolles, Fan, Florens, and Renault (2011) introduce Tikhonov regularization in the nonparametric IV model and obtain convergence rates under the source condition which quantifies the regularity of the estimated function relatively to the ill-posedness. Hall and Horowitz (2005) study the mildly ill-posed case and show optimality of obtained rates in the minimax sense. Chen and Reiss (2011) discuss connections between different sets of assumptions used in the nonparametric IV literature and obtain minimax rates for the severely ill-posed case under the very broad set of assumptions.

In the classical deconvolution problem, the density of the noise is known. Johannes et al. (2009) study the deconvolution problem, when the density of the noise is estimated from the additional sample, leading to the estimated convolution operator.

The functional linear regression model without endogeneity dates back at least to Cohen and Jones (1969) who suggest treating the problem with functional principal components. Hall, Horowitz, et al. (2007) obtain optimal rates for mildly-ill-posed case. Cardot and Johannes (2010) consider the orthogonal series estimator based on projection. Their set-up allows to obtain optimal rates for mildly and severely ill-posed problems, estimated derivatives, and the mean-squared prediction risk. The estimated operator in this class of models is a covariance operator.

The research on functional data in econometrics is very recent. Besides Florens and Van Bellegem (2015), another closely related study is Benatia, Carrasco, and Florens (2015). They consider the functional linear regression model with functional response and treat endogeneity with the functional instrumental variable. Another approach to high-dimensional econometric problems is to assume some sparsity and to use model selection devices, e.g. modifications of LASSO or Dantzig selector, see Belloni, Chernozhukov, and Hansen (2014), Gautier and Tsybakov (2014).

2 Identification

Suppose the relation between the scalar random variable Y and the random function $Z = (Z(t))_{t\in\mathscr{T}}$ is represented by the equation

$$Y = \int_{\mathscr{T}} \beta(t)Z(t)\mathrm{d}t + U,$$

where $\mathscr{T}$ is a compact subset of $\mathbf{R}$. Alternatively, if we start from the linear model $Y = \Phi(Z)+U$ with some continuous linear functional Φ, then by the Riesz representation theorem, $\Phi(Z) = \langle \beta, Z\rangle_{L^2_{\mathscr{T}}}$ with unique β.

Let the instrumental variable W has support $\mathcal{W} \subset \mathbf{R}^d$. The assumption $\mathbb{E}[UW] = 0$ is not sufficient to identify the infinite-dimensional object β. Consider the (generalized) Fourier expansion of β in some orthonormal basis of $L^2_{\mathscr{T}}$, $(\varphi_j)_{j\geq 1}$: $\beta = \sum_{k\geq 1}\langle \varphi_k, \beta\rangle_{L^2_{\mathscr{T}}}\varphi_k$. The model is transformed into the linear instrumental variable model with infinitely many endogenous regressors $(\langle Z, \varphi_j\rangle_{L^2_{\mathscr{T}}})_{j\geq 1}$ with generalized Fourier coefficients $(\langle \beta, \varphi_j\rangle_{L^2_{\mathscr{T}}})_{j\geq 1}$ being parameters of interest

$$Y = \sum_{j=1}^{\infty}\langle \beta, \varphi_j\rangle_{L^2_{\mathscr{T}}}\langle Z, \varphi_j\rangle_{L^2_{\mathscr{T}}} + U, \quad \mathbb{E}[UW] = 0.$$

If the function β is regular enough and the basis $(\varphi_j)_{j\geq 1}$ approximates it well, only first m coefficients are sufficient to represent the function β with a reasonable accuracy. Truncating the infinite sum, since $(\langle Z, \varphi_j\rangle_{L^2_{\mathscr{T}}})_{1\leq j\leq m}$ are correlated with U, one needs at least m moment conditions for the identification of m coefficients. The unconditional moment restriction does not identify β, unless $d \geq m$ instruments are available.

Strengthening the uncorrelatedness assumption to the mean-independence restriction, $\mathbb{E}[U|W] = 0$ a.s. with continuously distributed instrumental variable, however, gives us the infinite-number of moment conditions which can restore the identifying power.

Under suitable integrability conditions, $\mathbb{E}[U|W] = 0$ leads to

$$h(w) := \mathbb{E}[Y|W = w] = \int_{\mathscr{T}} \beta(t)\mathbb{E}[Z(t)|W = w]\mathrm{d}t =: (\mathcal{S}\beta)(w), \tag{2}$$

where $\mathcal{S} : L^2_{\mathscr{T}} \to L^2_W$ is an integral operator mapping a real square integrable function on $\mathscr{T}$ to another real function on $\mathcal{W}$, square integrable with respect to the law of W. Eq. (2) is an example of the Fredholm integral equation of type I solving which for $\beta(t)$ is known to be an ill-posed problem in most practical settings, Carrasco et al. (2007). The parameter β is identified when the operator $\mathcal{S}$ is injective. Alternatively, the null space of $\mathcal{S}$ should consists only from zero vector. Indeed, if this was not true, then $\mathcal{S}(\beta + \tilde{\beta}) = \mathcal{S}\beta + \mathcal{S}\tilde{\beta} = \mathcal{S}\beta, \forall\tilde{\beta} \in \mathcal{N}(\mathcal{S})$ and so we could not distinguish between $\beta + \tilde{\beta}$ and β. If $\mathcal{S}$ fails to be injective, the point identification is only possible in the orthogonal complement to the null space. Besides analytical interpretation, similarly to other models with conditional moment restrictions, the identifiability of β has also a statistical meaning in terms of the strength of the association between Z and W. We say that the $L^2_{\mathscr{T}}$-random element Z is *linearly complete* for W if for all $b \in L^2_{\mathscr{T}}$ such that $\mathbb{E}|\langle Z, b\rangle_{L^2_{\mathscr{T}}}| < \infty$, we have

$$\mathbb{E}\left[\langle Z, b\rangle_{L^2_{\mathscr{T}}}|W\right] = 0 \quad a.s. \implies b = 0 \quad a.e. \tag{3}$$

The linear completeness property of the distribution of (Z, W) is necessary and sufficient for the identification of β. This requirement is actually weaker than the natural generalization of completeness condition used in the nonparametric IV, Newey and Powell (2003). The natural generalization of completeness property to functional setting would also require Eq. (3) to hold for nonlinear functionals of Z. We make the following assumption on the data-generating process, to make sure that the model in Eq. (2) is well-defined.

Assumption 1.

1. *There exists β such that $\mathcal{S}\beta = h$, i.e., $h \in \mathcal{R}(\mathcal{S})$, where $\mathcal{R}(\mathcal{S})$ is the range of the operator $\mathcal{S}$.*
2. *The stochastic process Z is weakly complete for W.*

Consider another injective operator

$$\begin{aligned} \mathcal{U}:\ L^2_W &\to L^2_{\mathscr{S}} \\ \delta &\mapsto \mathbb{E}[\delta(W)\Psi(s,W)], \end{aligned}$$

where $\Psi(s,W), s \in \mathscr{S}$ is some integrable and measurable instrument function and $\mathscr{S}$ is a some subset of $\mathbf{R}^d$. Applying $\mathcal{U}$ to both sides of Eq. (2) gives

$$g(s) = \mathbb{E}[Y\Psi(s,W)] = \int_{\mathscr{T}} \beta(t)\mathbb{E}[Z(t)\Psi(s,W)]\mathrm{d}t = (\mathcal{T}\beta)(s), \tag{4}$$

where $g = \mathcal{U}h$ and $\mathcal{T} = \mathcal{U}\mathcal{S}$ is a new operator. It is more convenient to work with Eq. (4), since it does not involve conditional expectations which are estimated non-parametrically. Therefore, we get rid of the bandwidth selection problem and the curse of dimensionality.

Example 1. *Let $\delta \mapsto \int_{\mathcal{W}} \delta(w)\Psi(s,w)\mathrm{d}w$ be some non-singular integral operator with kernel function $\Psi(s,w)$. For instance if the support of IV is $[0,1]$, then it is not hard to verify that the integral operator with $\Psi(s,w) = (1+sw)e^{sw}$ is injective. Injectivity of $\mathcal{U}$ requires that $\mathcal{U}\delta = 0$ implies $\delta = 0$. Therefore, the density of the instrument should be bounded away from zero. It should also be such that $\delta f_W \in L^2_{[0,1]}, \forall \delta \in L^2_{[0,1]}$, which is the case, e.g. when it is continuous and bounded away from ∞,*

Example 2. *If $\Psi(s,W)$ is selected so that*

$$\mathbb{E}[U|W] = 0 \ a.s. \iff \mathbb{E}[U\Psi(s,W)] = 0, \quad \forall s \in \mathscr{S} \subset \mathbf{R}^d, \tag{5}$$

and so Eq. (2) holds if and only if Eq. (4) holds, the two sets of solutions coincide and we can ensure the unicity of β satisfying Eq. (4). This problem is also encountered in the consistent specification testing of the conditional mean function, Bierens (1982) and in the GMM-type estimators that exploit efficiently all information from conditional moment restrictions, see Dominguez and Lobato (2004); Lavergne and Patilea (2013). The following instrument functions Ψ enjoy the property in Eq. (5): $\mathbb{1}_{[W,+\infty)}(s), \exp(is^\top W), \exp(s^\top W), 1/(1+\exp(-s^\top W))$. (Stinchcombe and White, 1998, Corrolary 3.9.) characterize a large set of such functions. Namely, if $\Psi(s,W) = G(s^\top W)$ with $G : \mathbf{R} \to \mathbf{R}$ being real, analytic, but nonpolynomial, then Eq. (5) holds for any $\mathscr{S} \subset \mathbf{R}^d$ with positive Lebesgue measure. The class of such functions is very large and includes exponential, logarithmic, trigonometric functions and logistic CDF as a potential choices for G.

References

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. High-dimensional methods and inference on structural and treatment effects. *The Journal of Economic Perspectives*, 28(2): 29–50, 2014.

David Benatia, Marine Carrasco, and Jean-Pierre Florens. Functional linear regression with functional response. *Working paper*, 2015.

Herman J Bierens. Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134, 1982.

Hervé Cardot and Jan Johannes. Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, 101(2):395–408, 2010.

M. Carrasco, J.P. Florens, and E. Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6: 5633–5751, 2007.

Xiaohong Chen and Markus Reiss. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27(03):497–521, 2011.

Ayala Cohen and Richard H Jones. Regression on a random field. *Journal of the American Statistical Association*, 64(328):1172–1182, 1969.

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

Manuel A Dominguez and Ignacio N Lobato. Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72(5):1601–1615, 2004.

Jean-Pierre Florens and Sébastien Van Bellegem. Instrumental variable estimation in functional linear models. *Journal of Econometrics*, 186(2):465–476, 2015.

Eric Gautier and Alexandre B Tsybakov. High-dimensional instrumental variables regression and confidence sets. *arXiv preprint arXiv:1105.2454*, 2014.

Peter Hall and Joel L Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.

Peter Hall, Joel L Horowitz, et al. Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91, 2007.

Marc Hoffmann and Markus Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, pages 310–336, 2008.

Jan Johannes et al. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37 (5A):2301–2323, 2009.

Pascal Lavergne and Valentin Patilea. Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory. *Journal of Econometrics*, 177 (1):47–59, 2013.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

Maxwell B Stinchcombe and Halbert White. Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, 14(03):295–325, 1998.