On ill-posedness of nonparametric instrumental variable regression with convexity constraints

OLIVIER SCAILLET†

†University of Geneva and Swiss Finance Institute, Unimail, Bd du Pont d’Arve 40, CH-1211 Geneva 4, Switzerland.
E-mail: olivier скаilet@unige.ch

First version received: January 2016; final version accepted: August 2016

Summary This note shows that adding monotonicity or convexity constraints on the regression function does not restore well-posedness in nonparametric instrumental variable regression. The minimum distance problem without regularization is still locally ill-posed.

Keywords: Ill-posed inverse problems, Instrumental variable, Nonparametric estimation.

1. INTRODUCTION

We consider estimation of the regression model \( Y = \varphi_0(X) + U \). The variable \( X \) has compact support \( X = [0, 1] \) and is potentially endogenous. The instrument \( Z \) has compact support \( Z = [0, 1] \). The parameter of interest is the function \( \varphi_0 \) defined on \( X \), which satisfies the nonparametric instrumental variable regression (NPIVR):

\[
E[Y - \varphi_0(X)|Z] = 0.
\]

(1.1)

As shown in Example 1 of Gagliardini and Scaillet (2016), we do not need independence between the error \( U \) and the instrument \( Z \). This exemplifies a difference between restrictions induced by a parametric conditional moment setting and their nonparametric counterpart. NPIVR estimation has received considerable attention in recent years, building on a series of fundamental papers on ill-posed endogenous mean regressions; see Ai and Chen (2003), Newey and Powell (2003), Hall and Horowitz (2005), Blundell et al. (2007), Darolles et al. (2011), Horowitz (2011), and the review paper by Carrasco et al. (2007).

The main issue in nonparametric estimation with endogeneity is overcoming ill-posedness of the associated inverse problem. This occurs because the mapping of the reduced-form parameter (i.e. the distribution of the data) into the structural parameter (i.e. the instrumental regression function) is not continuous in the conditional moment \( E[Y|Z] \). We need a regularization of the estimation to recover consistency. Gagliardini and Scaillet (2012a) study a Tikhonov regularized (TiR) estimator; see Tikhonov (1963a,b), Groetsch (1984) and Kress (1999). They achieve regularization by adding a compactness-inducing penalty term, the Sobolev norm, to a functional minimum distance criterion. Chen and Pouzo (2012) study nonparametric estimation of conditional moment restrictions in which the generalized residual functions can be nonsmooth in the unknown functions of endogenous variables. For such a nonlinear NPIV problem, they propose a class of penalized sieve minimum distance estimators; see Chen and Pouzo (2015).
for inference in such a setting. As discussed in Matzkin (1994), in nonparametric models, we can use economic restrictions, as in parametric models, to reduce the variance of estimators, to falsify theories and to extrapolate beyond the support of the data. However, in addition, we can use some economic restrictions to guarantee the identification of some nonparametric models and the consistency of some nonparametric estimators. Economic theory often provides shape restrictions on functions of interest in applications, such as monotonicity, convexity and non-increasing (non-decreasing) returns to scale, but economic theory does not provide finite-dimensional parametric models. This motivates nonparametric estimation under shape restrictions. Because nonparametric estimates are often noisy, shape restrictions help to stabilize nonparametric estimates without imposing arbitrary restrictions; see the recent works of Blundell et al. (2012) and Horowitz and Lee (2015). Following that line of thought, we could hope that adding monotonicity or convexity constraints on the regression function would help to restore well-posedness in NPIVR.

The next section shows that this is unfortunately not the case as the minimum distance problem without regularization is still locally ill-posed. Chetverikov and Wilhelm (2015) look at imposing two monotonicity conditions: (a) monotonicity of the regression function \( \phi_0 \) and (b) monotonicity of the reduced-form relationship between the endogenous regressor \( X \) and the instrument \( Z \), in the sense that the conditional distribution of \( X \) given \( Z \) corresponding to higher values of \( Z \) first-order stochastically dominates the same conditional distribution corresponding to lower values of \( Z \) (the monotonic IV assumption). They show that these two monotonicity conditions together significantly change the structure of the NPIV model, and weaken its ill-posedness. In particular, they point out that, even if well-posedness is not restored, those two monotonicity constraints improve the rate of convergence in shrinking neighbourhoods of the constraint boundary and can have a significant impact on the estimator finite sample behaviour. Chen and Christensen (2013) show that imposing shape restrictions only is not enough to improve convergence rates as long as the derivative constraints hold with strict inequality (i.e. in the interior of the constraint space). There may be rate improvements when the constraint is binding.

2. ILL-POSEDNESS WITH CONVEXITY CONSTRAINTS

The functional parameter \( \phi_0 \) belongs to a subset \( \Theta \) of \( L^2(X) \), where \( L^2(X) \) denotes the \( L^2 \) space of square integrable functions of \( X \) defined by the scalar product \( \langle \phi, \psi \rangle = \int \phi(x)\psi(x)dx \), and we write \( ||\phi|| \) for the \( L^2 \) norm \( \langle \phi, \phi \rangle^{1/2} \).

We assume the following identification condition.

**Assumption 2.1.** \( \phi_0 \) is the unique function \( \varphi \in L^2(X) \) that satisfies the conditional moment restriction (1.1).

We refer to Newey and Powell (2003), Theorems 2.2–2.4, for sufficient conditions ensuring Assumption 2.1.

Let us consider a nonparametric minimum distance approach to obtain \( \varphi_0 \). This relies on \( \varphi_0 \) minimizing

\[
Q_\infty(\varphi) := E[m(\varphi, Z)^2], \quad \varphi \in L^2(X),
\]

where \( m(\varphi, Z) = E[Y - \varphi(X)|Z] \). We can write the conditional moment function \( m(\varphi, z) \) as

\[
m(\varphi, z) = (A\varphi)(z) - r(z) = (A\Delta\varphi)(z),
\]

© 2016 Royal Economic Society.
with \( \Delta \varphi := \varphi - \varphi_0 \). Here, the linear operator \( A \) is defined by

\[
(A \varphi)(z) := \int \varphi(x)f_{X|Z}(x|z)dx + \int yf_{Y|Z}(y|z)dy,
\]

where \( f_{X|Z} \) and \( f_{Y|Z} \) are the conditional densities of \( X \) given \( Z \), and \( Y \) given \( Z \). Assumption 2.1 on identification of \( \varphi_0 \) holds if and only if operator \( A \) is injective. Further, we assume that \( A \) is a bounded operator from \( L_2(\mathcal{X}) \) to \( L_2(\mathcal{Z}) \), where \( L_2(Z) \) denotes the \( L_2 \) space of square integrable functions of \( Z \) defined by the scalar product

\[
\langle \psi_1, \psi_2 \rangle_{L_2(Z)} := E[\psi_1(Z)\psi_2(Z)].
\]

The limit criterion (2.1) becomes

\[
Q_\infty(\varphi) = \langle A \Delta \varphi, A \Delta \varphi \rangle_{L_2(Z)}. \tag{2.3}
\]  

**ASSUMPTION 2.2.** The linear operator \( A \) from \( L_2(\mathcal{X}) \) to \( L_2(\mathcal{Z}) \) is compact.

Assumption 2.2 on compactness of operator \( A \) holds under mild conditions on the conditional density \( f_{X|Z} \); see, e.g., Gagliardini and Scaillet (2012a). In the proof of Proposition 2.1 in the Appendix, we also need the regularity conditions: \( \sup_z |f_Z(z)| < \infty \) and \( \sup_{x,z} |f_{X|Z}(x|z)| < \infty \).

Proposition 2.1 shows that the minimum distance problem above is locally ill-posed – see, e.g. Definition 1.1 in Hofmann and Scherzer (1998) – even if we consider monotonicity, monotonicity nonnegativity or convexity constraints. There are sequences of increasingly oscillatory functions arbitrarily close to \( \varphi_0 \) that approximately minimize \( Q_\infty \) while not converging to \( \varphi_0 \). In other words, function \( \varphi_0 \) is not identified in \( \Theta \) as an isolated minimum of \( Q_\infty \). Therefore, ill-posedness can lead to inconsistency of the naïve analogue estimators based on the empirical analogue of \( Q_\infty \). In order to rule out these explosive solutions, we can use penalization as in Gagliardini and Scaillet (2012a); see Gagliardini and Scaillet (2012b) and Chen and Pouzo (2012) for the quantile regression case. Under a stronger assumption than Assumption 2.1, namely local injectivity of \( A \), the definition of local ill-posedness is equivalent to \( A^{-1} \) being discontinuous in a neighbourhood of \( A(\varphi_0) \); see Chapter 10 of Engl et al. (2000).

**PROPOSITION 2.1.** Let \( \varphi_0 \) satisfy monotonicity, monotonicity nonnegativity or convexity constraints. Then, under Assumptions 2.1 and 2.2, the minimum distance problem is locally ill-posed, namely for any \( r > 0 \) small enough, there exist \( \varepsilon \in (0, r) \) and a sequence \( \{\varphi_n\} \subset B_r(\varphi_0) := \{\varphi \in L_2(\mathcal{X}) : \|\varphi - \varphi_0\| < r\} \) such that \( \|\varphi_n - \varphi_0\| \geq \varepsilon \), \( Q_\infty(\varphi_n) \to Q_\infty(\varphi_0) = 0 \), and such that \( \varphi_n \) satisfies the same constraints as \( \varphi_0 \).

**ACKNOWLEDGEMENTS**

We thank Xiaohong Chen and Daniel Wilhelm for helpful comments.

**REFERENCES**

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–843.

Blundell, R., X. Chen and D. Kristensen (2007). Semi-nonparametric IV estimation of shape invariant Engel curves. *Econometrica* 75, 1613–69.

Blundell, R., J. Horowitz and M. Parey (2012). Measuring the price responsiveness of gasoline demand: economic shape restrictions and nonparametric demand estimation. *Quantitative Economics* 3, 29–51.

Carrasco, M., J.-P. Florens and E. Renault (2007). Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume 6*, 5633–751. Amsterdam: North-Holland.

© 2016 Royal Economic Society.
On ill-posedness

Chen, X. and T. Christensen (2013). Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. CWP 56/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.

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

Chen, X. and D. Pouzo (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* 83, 1013–79.

Chetverikov, D. and D. Wilhelm (2015). Nonparametric instrumental variable estimation under monotonicity. CWP 39/15, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.

Darolles, S., Y. Fan, J.-P. Florens and E. Renault (2011). Nonparametric instrumental regression. *Econometrica* 79, 1541–65.

Engl, H., M. Hanke and A. Neubauer (2000). *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic.

Gagliardini, P. and O. Scaillet (2012a). Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics* 167, 61–75.

Gagliardini, P. and O. Scaillet (2012b). Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica* 80, 1533–62.

Gagliardini, P. and O. Scaillet (2016). A specification test for nonparametric instrumental variable regression. Forthcoming in *Annals of Economics and Statistics, Special Issue on Inverse Problems in Econometrics*.

Groetsch, C. (1984). *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Boston: Pitman Advanced Publishing Program.

Hall, P. and J. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* 33, 2904–29.

Hofmann, B. and O. Scherzer (1998). Local ill-posedness and source conditions of operator equations in Hilbert spaces. *Inverse Problems* 14, 1189–206.

Horowitz, J. (2011). Applied nonparametric instrumental variables estimation. *Econometrica* 79, 347–94.

Horowitz, J. and S. Lee (2015). Nonparametric estimation and inference under shape restrictions. CWP 67/15, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.

Kress, R. (1999). *Linear Integral Equations*. New York: Springer.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 6*, 2523–58. Amsterdam: North-Holland.

Newey, W. and J. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–78.

Tikhonov, A. (1963a). On the solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady* 4, 1035–8 (English translation).

Tikhonov, A. (1963b). Regularization of incorrectly posed problems. *Soviet Mathematics Doklady* 4, 1624–7 (English translation).

**APPENDIX: PROOFS OF RESULTS**

**Proof of Proposition 2.1:** The proof gives explicit sequences \((\varphi_n)\) generating ill-posedness when \(\varphi_0\) satisfies monotonicity, monotonicity nonnegativity or convexity constraints.
Let us build $\varphi_n = \varphi_0 + \varepsilon \psi_n$, $\varepsilon > 0$, where $\psi_n(x) := -(2n + 1)^{1/2}(1 - x)^n$ and $\varphi_0$ is monotonic and increasing. Then $\varphi_n \in L^2(X)$ and the first derivative $\nabla \varphi_n \geq 0$. Because $\|\psi_n\| = 1$, when we choose $\varepsilon > 0$ sufficiently small, we have $(\varphi_n) \subset B_r(\varphi_0)$, and $\varphi_n \nrightarrow \varphi_0$. We also have that $A\varphi_n \nrightarrow A\varphi_0$, where $\nrightarrow$ denotes weak convergence. Indeed, for $q \in L^2(\mathbb{Z})$, we obtain $(q, A\varphi_n)_{L^2(\mathbb{Z})} = (q, A\varphi_0)_{L^2(\mathbb{Z})} + \varepsilon (q, A\psi_n)_{L^2(\mathbb{Z})}$, and $(q, A\psi_n)_{L^2(\mathbb{Z})} \rightarrow 0$, as $|A\psi_n| \leq C(2n + 1)^{1/2}(1/(n + 1))$ for $C > 0$. As $A$ is compact and $(\varphi_n)$ is bounded, the sequence $A\varphi_n$ admits a convergent subsequence $A\varphi_m(n) \rightarrow \xi$. Because the weak limit is unique, we have $\xi = A\varphi_0$. Thus, $A\varphi_m(n) \rightarrow A\varphi_0$ and $Q_\infty(\varphi_m(n)) \rightarrow 0$ but $\|\varphi_m(n) - \varphi_0\| \geq \varepsilon$, and hence the stated result follows.

The above argument works also with the function

$$
\psi_n(x) := \frac{(2n + 1)^{1/2}}{(2n + 1)} (1 + x)^n,
$$

which yields a monotonic nonnegative function $\varphi_n \in L^2(X)$ if $\varphi_0 \geq 0$ and is monotonic. This shows that the positivity constraint does not help here either.

As the higher-order derivatives $\nabla^m \psi_n \geq 0$, $m \geq 1$, this example also shows that positivity constraints on the higher-order derivatives $\nabla^m \varphi_0 \geq 0$, such as a convexity constraint $\nabla^2 \varphi_0 \geq 0$, does not restore well-posedness of the estimation problem in the NPIVR setting. \hfill \Box