

Decentralization estimators for instrumental variable quantile regression models

HIROAKI KAIDO

Department of Economics, Boston University

KASPAR WÜTHRICH

Department of Economics, University of California San Diego

The instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen (2005)) is a popular tool for estimating causal quantile effects with endogenous covariates. However, estimation is complicated by the nonsmoothness and nonconvexity of the IVQR GMM objective function. This paper shows that the IVQR estimation problem can be decomposed into a set of conventional quantile regression subproblems which are convex and can be solved efficiently. This reformulation leads to new identification results and to fast, easy to implement, and tuning-free estimators that do not require the availability of high-level “black box” optimization routines.

KEYWORDS. Instrumental variables, quantile regression, contraction mapping, fixed-point estimator, bootstrap.

JEL CLASSIFICATION. C21, C26.

1. INTRODUCTION

Quantile regression (QR), introduced by Koenker and Bassett (1978), is a widely-used method for estimating the effect of regressors on the whole outcome distribution. QR is flexible, easy to interpret, and can be computed very efficiently as the solution to a convex problem. In many applications, the regressors of interest are endogenous, rendering QR inconsistent for estimating causal quantile effects. The instrumental variable quantile regression (IVQR) model of Chernozhukov and Hansen (2005, 2006) generalizes QR to accommodate endogenous regressors. Unfortunately, in sharp contrast to QR and other IV estimators such as two-stage least squares (2SLS), estimation of IVQR models is

Hiroaki Kaido: hkaido@bu.edu

Kaspar Wüthrich: kwuthrich@ucsd.edu

We are grateful to Alfred Galichon for many helpful discussions and valuable inputs. We would like to thank Victor Chernozhukov, Patrik Guggenberger, Hidehiko Ichimura, Yuichiro Kamada, David Kaplan, Hiroyuki Kasahara, Kengo Kato, Ivana Komunjer, Blaise Melly, Klaus Neusser, Quang Vuong (discussant), anonymous referees, and conference and seminar participants at Boston College, Columbia University, Penn State, the University of Arizona, the University of Bern, UCLA, UCR, UCSB, UCSD, UNC, the Interactions: Bringing Together Econometrics and Applied Microeconomics Conference 2017, and the Workshop for Advances in Microeconometrics 2017 for helpful comments. We are grateful to Chris Hansen for sharing the data for the empirical application. This project was started when both authors were visiting the MIT in 2015. Wüthrich is also affiliated with CESifo and ifo Institute.

computationally challenging because the resulting estimation problem, formulated as a generalized method of moments (GMM) problem, is nonsmooth and nonconvex. From an applied perspective, this issue is particularly troublesome as resampling methods are often used to avoid the choice of tuning parameters when estimating the asymptotic variance of IVQR estimators.

In this paper, we develop a new class of estimators for linear IVQR models. The proposed estimators are fast, easy to implement, tuning-free, and do not require the availability of high-level “black box” optimization routines. They are particularly suitable for settings with many exogenous regressors, a moderate number of endogenous regressors, and a large number of observations, which are ubiquitous in applied research. The key insight underlying our estimators is that the complicated and nonlinear IVQR estimation problem can be “decentralized,” that is, decomposed into a set of more tractable subproblems, each of which is solved by a “player” who best responds to the other players’ actions. Each subproblem is a conventional (weighted) QR problem, which is convex and can be solved very quickly using robust algorithms. The IVQR estimator is then characterized as a fixed point of such subproblems, which can be viewed as the pure strategy Nash equilibrium of the “game.” Computationally, this reformulation allows us to recast the original nonsmooth and nonconvex optimization problem as the problem of finding the fixed point of a low dimensional map, which leads to substantial reductions in computation times.

Implementation of our preferred procedures is straightforward and only requires the availability of a routine for estimating QRs and in some cases a univariate root-finder. The resulting estimation algorithms attain significant computational gains. For example, we show that in problems with two endogenous variables, a version of our estimator that uses a contraction algorithm is 149–308 times faster than the most popular existing approach for estimating IVQR models, the inverse quantile regression (IQR) estimator of [Chernozhukov and Hansen \(2006\)](#). Another version that uses a nested root-finding algorithm, which is guaranteed to converge under a milder condition, is 84–134 times faster than the IQR estimator. Importantly, these computational gains do not come at a cost in terms of the finite sample performance of our procedures, which is very similar to IQR. The computational advantages of our estimators are even more substantial with more than two endogenous variables. The reason is that the dimensionality of the grid search underlying IQR corresponds to the number of endogenous variables, which renders IQR computationally prohibitive in empirically relevant settings whenever the number of endogenous variables exceeds two or three.

The fixed-point reformulation also provides new insights into global identification of IVQR models. In particular, it allows us to study identification and stability of the algorithms (at the population level) in the same framework. Exploiting the equivalence of global identification and uniqueness of the fixed point, we give a new identification result and population algorithms based on the contraction mapping theorem. We then compare our identification conditions to those of [Chernozhukov and Hansen \(2006\)](#). Further, our reformulation is shown to be useful beyond setups where the contraction mapping theorem applies as long as the parameter of interest is globally identified. For such settings, algorithms based on root-finding methods are proposed. Finally, we show

that, by recursively nesting fixed-point problems, it is possible to recast the IVQR estimation problem as a univariate root-finding problem. While adding nests incurs additional computational costs, our Monte Carlo experiments suggest that estimation procedures based on nesting perform well and are computationally reasonable when the number of endogenous variables is moderate.

We establish consistency and asymptotic normality of the proposed estimators and prove validity of the empirical bootstrap for estimating the limiting laws. The bootstrap is particularly attractive in conjunction with our computationally efficient estimation algorithms as it allows us to avoid the choice of tuning parameters inherent to estimating the asymptotic variance based on analytic formulas. The key technical ingredient for deriving our theoretical results is the Hadamard differentiability of the fixed-point map. This result may be of independent interest.

To illustrate the usefulness of our estimation algorithms, we revisit the analysis of the impact of 401(k) plans on savings in [Chernozhukov and Hansen \(2004\)](#). Based on this application, we perform extensive Monte Carlo simulations, which demonstrate that our estimation and inference procedures have excellent finite sample properties.

1.1 Literature

We contribute to the literature on estimation and inference based on linear IVQR models. [Chernozhukov and Hong \(2003\)](#) have proposed a quasi-Bayesian approach which can accommodate multiple endogenous variables but, as noted by [Chernozhukov and Hansen \(2013\)](#), “requires careful tuning in applications” (p. 75). [Chernozhukov and Hansen \(2006\)](#) have developed an inverse QR algorithm that combines grid search with convex QR problems. Because the dimensionality of the grid search equals the number of endogenous variables, this approach is computationally feasible only if the number of endogenous variables is very low. [Chernozhukov and Hansen \(2008\)](#) and [Jun \(2008\)](#) have studied weak instrument robust inference procedures based on the inversion of Anderson–Rubin-type tests. [Chernozhukov, Hansen, and Jansson \(2009\)](#) have proposed a finite sample inference approach. [Andrews and Mikusheva \(2016\)](#) have developed a general conditional inference approach and derived sufficient conditions for the IVQR model. [Kaplan and Sun \(2017\)](#) and [de Castro, Galvao, Kaplan, and Liu \(2019\)](#) have suggested using smoothed estimating equations to overcome the non-smoothness of the IVQR estimation problem; however, the nonconvexity persists. More recently, [Chen and Lee \(2018\)](#) have proposed to recast the IVQR problem as a mixed-integer quadratic programming problem that can be solved using well-established algorithms. However, efficiently solving such a problem is still challenging even for low-dimensional settings. [Zhu \(2018\)](#) has shown that if the ℓ_∞ norm is used rather than the ℓ_2 norm, the problem admits a reformulation as a mixed-integer linear programming problem, which can be solved more efficiently than the quadratic program in [Chen and Lee \(2018\)](#). This procedure typically requires an early termination of the algorithm to ensure computational tractability which is akin to a tuning parameter choice. In addition, [Zhu \(2018\)](#) has proposed a k -step approach that allows for estimating models with multiple endogenous regressors based on large datasets, but requires estimating the gradient. [Pouliot](#)

(2019) proposed a mixed integer linear programming formulation that allows for sub-vector inference via the inversion of a distribution-free rankscore test and can be modified to accommodate weak instruments. An important drawback of the estimation approaches based on mixed integer reformulations is that they rely on the availability of high-level “black box” optimization routines such as Gurobi and often require careful tuning in applications. Finally, imposing a location-scale model for the potential outcomes, Machado and Santos Silva (2019) proposed moment-based estimators for the structural quantile function.

Compared to the existing literature on the estimation of linear IVQR models, the main advantages of the proposed estimation algorithms are the following. First, by relying on convex QR problems, our estimators are easy to implement, robust, and computationally efficient in settings with many exogenous variables, a moderate number of endogenous variables, and a large number of observations. Second, by exploiting the specific structure of the IVQR estimation problem, our estimators are tuning-free and do not require the availability of high-level “black box” optimization routines. Third, our estimators are based on the original IVQR estimation problem, and thus avoid the choice of smoothing bandwidths and do not rely on additional restrictions on the structural quantile function.

Semi- and nonparametric estimation of IVQR models have been considered by Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007), Chen and Pouzo (2009), Chen and Pouzo (2012), Gagliardini and Scaillet (2012), and Wüthrich (2019). We refer to Chernozhukov and Hansen (2013) and Chernozhukov, Hansen, and Wüthrich (2017) for reviews of the IVQR model and references to empirical applications.

Abadie, Angrist, and Imbens (2002) have proposed an alternative approach to the identification and estimation of quantile effects with binary endogenous regressors, which builds on the local average treatment effects framework of Imbens and Angrist (1994). Their approach has been extended and further developed by Frandsen, Frölich, and Melly (2012), Frölich and Melly (2013), and Belloni, Chernozhukov, Fernandez-Val, and Hansen (2017) among others. We refer to Melly and Wüthrich (2017) for a recent review of this approach and to Wüthrich (2020) for a comparison between this approach and the IVQR model. Identification and estimation of nonseparable models with continuous endogenous regressors have been studied by Chesher (2003), Ma and Koenker (2006), Lee (2007), Jun (2009), Imbens and Newey (2009), D’Haultfoeuille and Février (2015), and Torgovitsky (2015) among others.

On a broader level, our paper contributes to the literature, which proposes estimation procedures that rely on decomposing computationally burdensome estimation problems into several more tractable subproblems. This type of procedure, which we call decentralization, has been applied in different contexts. Examples include the estimation of single index models with unknown link function (Weisberg and Welsh (1994)), general maximum likelihood problems (Smyth (1996)), linear models with high-dimensional fixed effects (e.g., Guimaraes and Portugal (2010)), and the references therein, sample selection models (Marra and Radice (2013)), peer effects models (Arcidiacono, Foster, Goodpaster, and Kinsler (2012)), interactive fixed effects models (e.g.,

Chen, Fernandez-Val, and Weidner (2014), Moon and Weidner (2015)), and random coefficient logit demand models (Lee and Seo (2015)). Most of these papers decompose a single estimation problem into two subproblems. The present paper explicitly considers cases in which the number of subproblems may exceed two.

1.2 Organization of the paper

The remainder of the paper is structured as follows. Section 2 introduces the setup and the IVQR model. Section 3 shows that the IVQR estimation problem can be decentralized into a series of (weighted) conventional QR problems. In Section 4, we introduce population algorithms based on the contraction mapping theorem and root-finders. Section 5 discusses the corresponding sample algorithms. In Section 6, we establish the asymptotic normality of our estimators and prove the validity of the bootstrap. Section 7 presents an empirical application. In Section 8, we provide simulation evidence on the computational performance and the finite sample properties of our methods. Section 9 concludes. All proofs as well as some additional theoretical and simulation results are collected in the Appendix in the Online Supplementary Material (Kaido and Wüthrich (2021)).

2. SETUP AND MODEL

Consider a setup with a continuous outcome variable Y , a $d_X \times 1$ vector of exogenous covariates X , a $d_D \times 1$ vector of endogenous treatment variables D , and a $d_Z \times 1$ vector of instruments Z . The IVQR model is developed within the standard potential outcomes framework (e.g., Rubin (1974)). Let $\{Y_d\}$ denote the (latent) potential outcomes. The object of primary interest is the conditional quantile function of the potential outcomes, which we denote by $q(d, x, \tau)$. Having conditioned on covariates $X = x$, by the Skorohod representation of random variables, potential outcomes can be represented as

$$Y_d = q(d, x, U_d) \quad \text{with } U_d \sim \text{Uniform}(0, 1).$$

This representation lies at the heart of the IVQR model. With this notation at hand, we state the main conditions of the IVQR model (Chernozhukov and Hansen (2005, Assumptions A1–A5)).

ASSUMPTION 1. *Given a common probability space (Ω, F, P) , the following conditions hold jointly with probability one:*

- (1) *Potential outcomes: Conditional on $X = x$, for each d , $Y_d = q(d, x, U_d)$, where $q(d, x, \tau)$ is strictly increasing in τ and $U_d \sim \text{Uniform}(0, 1)$.*
- (2) *Independence: Conditional on $X = x$, $\{U_d\}$ are independent of Z .*
- (3) *Selection: $D := \delta(Z, X, V)$ for some unknown function $\delta(\cdot)$ and random vector V .*
- (4) *Rank invariance or rank similarity: Conditional on $X = x$, $Z = z$,*

- (a) $\{U_d\}$ are equal to each other; or, more generally,
- (b) $\{U_d\}$ are identically distributed, conditional on V .

(5) *Observed variables:* Observed variables consist of $Y := q(D, X, U_D)$, D , X , and Z .

We briefly discuss the most important aspects of Assumption 1 and refer the interested reader to Chernozhukov and Hansen (2005, 2006, 2013) for more comprehensive treatments. Assumption 1(1) states the Skorohod representation of Y_d and requires strict monotonicity of the potential outcome quantile function, which rules out discrete outcomes. Assumption 1(2) imposes independence between potential outcomes and instruments. Assumption 1(3) defines a general selection mechanism. The key restriction of the IVQR model is Assumption 1(4). Rank invariance (a) requires individual ranks U_d to be the same across treatment states. Rank similarity (b) weakens this condition, allowing for random slippages of U_d away from a common level U . Finally, Assumption 1(5) summarizes the observables.

REMARK 2.1. Assumption 1 does not impose any restrictions on how the instrument Z affects the endogenous variable D . As a consequence, Assumption 1 alone does not guarantee point identification of the structural quantile function. In Sections 3 and 4, we discuss sufficient conditions for global (point) identification. Due to the nonlinearity of the IVQR problem, these conditions are stronger than the usual first stage assumptions in linear instrumental variables models and require the instrument to have a nontrivial impact on the joint distribution of (Y, D) . The identification conditions are particularly easy to interpret when D and Z are binary; see Chernozhukov and Hansen (2005, Section 2.4) and Appendix C.2.1 for a further discussion.

The main implication of Assumption 1 is the following conditional moment restriction (Chernozhukov and Hansen (2005, Theorem 1)):

$$P(Y \leq q(D, X, \tau) \mid X, Z) = \tau, \quad \tau \in (0, 1). \quad (2.1)$$

In this paper, we focus on the commonly used linear-in-parameters model for $q(\cdot)$ (e.g., Chernozhukov and Hansen (2006)):

$$q(d, x, \tau) = x' \theta_X(\tau) + d' \theta_D(\tau),$$

where $\theta(\tau) := (\theta_X(\tau)', \theta_D(\tau)')' \in \mathbb{R}^{d_X + d_D}$ is the finite-dimensional parameter vector of interest. The conditional moment restriction (2.1) suggests GMM estimators based on the following unconditional population moment conditions:

$$\Psi_P(\theta(\tau)) := E_P \left[(1\{Y \leq X' \theta_X(\tau) + D' \theta_D(\tau)\} - \tau) \begin{pmatrix} X \\ Z \end{pmatrix} \right].$$

Our primary goal here is to obtain estimators in a computationally efficient and reliable manner. We therefore focus on just-identified moment restrictions where $d_Z = d_D$,

for which the construction of an estimator is straightforward. A potential caveat of this approach is that estimators based on these restrictions do not achieve the pointwise (in τ) semiparametric efficiency bound implied by the conditional moment restrictions (2.1). Appendix A provides a discussion of overidentified GMM problems and presents a two-step approach for constructing efficient estimators based on the proposed algorithms.

In what follows, we will often suppress the dependence on τ to lighten-up the exposition. We then define the true parameter value θ^* as the solution to the moment conditions, that is,

$$\Psi_P(\theta^*) = 0.$$

The resulting GMM objective function reads

$$Q_N^{\text{GMM}}(\theta) = -\frac{1}{2} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N m_i(\theta) \right)' W_N(\theta) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N m_i(\theta) \right), \tag{2.2}$$

where $m_i(\theta) := (1\{Y_i \leq X_i'\theta_X + D_i'\theta_D\} - \tau)(X_i', Z_i)'$ and $W_N(\theta)$ is a positive definite weighting matrix. Estimation based on (2.2) is complicated by the nonsmoothness and, most importantly, the nonconvexity of Q_N^{GMM} . This paper proposes a new set of algorithms to address these challenges.

3. DECENTRALIZATION

Here, we describe the basic idea behind our decentralization estimators. To simplify the exposition, we first illustrate our approach with the population problem of finding the true parameter value θ^* in the IVQR model. Our estimator then adopts the analogy principle, which will be presented in Section 5. The key insight is that the complicated nonlinear IVQR estimation problem can be “decentralized,” that is, decomposed into a set of more tractable subproblems, each of which is solved by a “player” who best responds to other players’ actions. Specifically, we first split the parameter vector θ into J subvectors $\theta_1, \dots, \theta_J$, where $J = d_D + 1$. We then decompose the grand estimation problem into J subproblems. Each of the subproblems is allocated to a distinct player. For each j , player j ’s choice variable is the j th subvector θ_j . Her problem is to find the value of θ_j such that a subset of the moment restrictions is satisfied given the other players’ actions θ_{-j} , where θ_{-j} stacks the components of θ other than θ_j . This reformulation allows us to view the estimation problem as a game of complete information and to characterize θ^* as the game’s pure strategy Nash equilibrium.

We start by describing the parameter subvectors. First, let $\theta_1 = \theta_X \in \mathbb{R}^{d_X}$ denote the vector of coefficients on the exogenous variables. We allocate this subvector to player 1. Similarly, for each $j = 2, \dots, J$, let $\theta_j \in \mathbb{R}$ denote the coefficient on the $(j - 1)$ -th endogenous variable, which is allocated to player j . The coefficient vector for the endogenous variable can therefore be written as $\theta_D = (\theta_2, \dots, \theta_J)'$. For each $\theta \in \mathbb{R}^{d_X + d_D}$, define the following (weighted) QR objective functions:

$$Q_{P,1}(\theta) := E_P[\rho_\tau(Y - X'\theta_1 - D_1\theta_2 - \dots - D_{d_D}\theta_J)], \tag{3.1}$$

$$Q_{P,j}(\theta) := E_P[\rho_\tau(Y - X'\theta_1 - D_1\theta_2 - \dots - D_{d_D}\theta_J)(Z_{j-1}/D_{j-1})], \quad j = 2, \dots, J, \tag{3.2}$$

where $\rho_\tau(u) = u(\tau - 1\{u < 0\})$ is the “check-function.” We assume that the model is parametrized such that Z_ℓ/D_ℓ is positive for all $\ell = 1, \dots, d_D$. Under our assumptions, we can always reparametrize the model such that this condition is met; see Appendix B for more details.

The players then solve the following optimization problems:

$$\begin{aligned} \min_{\tilde{\theta}_1 \in \mathbb{R}^{d_X}} Q_{P,1}(\tilde{\theta}_1, \theta_{-1}), \\ \min_{\tilde{\theta}_j \in \mathbb{R}} Q_{P,j}(\tilde{\theta}_j, \theta_{-j}), \quad j = 2, \dots, J. \end{aligned}$$

Observe that each player’s problem is a weighted QR problem, which is convex in its choice variable. For the sample analogues of these problems fast solution algorithms exist (e.g., [Koenker \(2017\)](#)).

For each j , let $\tilde{L}_j(\theta_{-j})$ denote the set of minimizers. Borrowing the terminology from game theory, we refer to these maps as *best response (BR) maps*. Under an assumption, we specify below the first-order optimality conditions imply that, for each j , any element $\tilde{\theta}_j^* \in \tilde{L}_j(\theta_{-j})$ of the BR map satisfies

$$\Psi_{P,1}(\tilde{\theta}_1^*, \theta_{-1}) := E_P[(1\{Y \leq X'\tilde{\theta}_1^* + D'\theta_{-1}\} - \tau)X] = 0, \quad (3.3)$$

$$\begin{aligned} \Psi_{P,j}(\tilde{\theta}_j^*, \theta_{-j}) &:= E_P[(1\{Y \leq (X', D'_{-(j-1)})'\theta_{-j} + D_{j-1}\tilde{\theta}_j^*\} - \tau)Z_{j-1}] \\ &= 0, \quad j = 2, \dots, J, \end{aligned} \quad (3.4)$$

where $D_{-(j-1)}$ stacks a vector all endogenous variables except D_{j-1} . Note that $\Psi_P(\theta) = (\Psi_{P,1}(\theta)', \dots, \Psi_{P,J}(\theta)')$ is the set of unconditional IVQR moment conditions. Hence, θ^* satisfies

$$\theta_j^* \in \tilde{L}_j(\theta_{-j}^*), \quad j = 1, \dots, J,$$

which implies that θ^* is a fixed point of the BR-maps (i.e., a Nash equilibrium of the game).

In what follows, we work with conditions that ensure the existence of singleton-valued BR maps L_j , $j = 1, \dots, J$, such that, for each j , $\Psi_{P,j}(L_j(\theta_{-j}), \theta_{-j}) = 0$.¹ We say that the IVQR estimation problem admits *decentralization* if there exist such BR functions defined over domains for which the moment conditions can be evaluated.² To ensure decentralization, we make the following assumption.

ASSUMPTION 2. *The following conditions hold:*

- (1) Θ is a closed rectangle in $\mathbb{R}^{d_X+d_D}$. θ^* is in the interior of Θ .

¹While it may be interesting to work with set-valued maps, the existence of the BR functions greatly simplifies our analysis of identification and inference.

²In Appendix C.2, we also provide weaker conditions under which the decentralization holds on a local neighborhood of θ^* . We call such a result *local decentralization*, which is sufficient for analyzing the (local) asymptotic behavior of the estimator.

- (2) $E[|Z_\ell|^2] < \infty$ for $\ell = 1, \dots, d_D$. $E[|X_k|^2] < \infty$ for all $k = 1, \dots, d_X$. For each $\ell = 1, \dots, d_D$, D_ℓ has a compact support;
- (3) The conditional cdf $y \rightarrow F_{Y|D,X,Z}(y)$ is continuously differentiable for all $y \in \mathbb{R}$ a.s. The conditional density $f_{Y|D,Z,X}$ is uniformly bounded a.s.;
- (4) For any $\theta \in \Theta$, the matrices

$$E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)XX']$$

and

$$E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)D_\ell Z_\ell], \quad \ell = 1, \dots, d_D,$$

are positive definite.

For each j , let $\Theta_{-j} \subset \mathbb{R}^{d-j}$ denote the parameter space for θ_{-j} . Assumption 2(1) ensures that Θ is compact. This assumption also ensures that each Θ_{-j} is a closed rectangle, which we use to show that L_j is well-defined on a suitable domain. Assumptions 2(2) and 2(3) impose standard regularity conditions on the conditional density and the moments of the variables in the model. We assume D_ℓ has a compact support, which allows us to always reparametrize the model so that the objective function in (3.2) is well-defined and convex (cf. Appendix B). The first part of Assumption 2(4) is a standard full rank condition which is a natural extension of the local full rank condition required for local identification and decentralization (cf. Assumption 4 in the Appendix). For the second part of Assumption 2(4), it suffices that the model is parametrized such that, for each $\ell \in \{1, \dots, d_D\}$, $D_\ell Z_\ell$ (and Z_ℓ/D_ℓ) is positive with probability one.

For each j , define

$$R_{-j} := \{\theta_{-j} \in \Theta_{-j} : \Psi_{P,j}(\theta) = 0, \text{ for some } \theta = (\theta_j, \theta_{-j}) \in \Theta\}. \quad (3.5)$$

This is the set of subvectors θ_{-j} for which one can find $\theta_j \in \Theta_j$ such that $\theta = (\theta_j, \theta_{-j})'$ solve the j th moment restriction. We take this set as the domain of player j 's best response function L_j .

The following lemma establishes that the IVQR model admits decentralization.

LEMMA 1. *Suppose that Assumptions 1 and 2 hold. Then, there exist functions $L_j : R_{-j} \rightarrow \mathbb{R}^{d_j}$, $j = 1, \dots, J$ such that, for $j = 1, \dots, J$,*

$$\Psi_{P,j}(L_j(\theta_{-j}), \theta_{-j}) = 0, \quad \text{for all } \theta_{-j} \in R_{-j}.$$

Further, L_j is continuously differentiable on the interior of R_{-j} for all $j = 1, \dots, J$.

We now introduce maps that represent all players' (joint) best responses. We consider two basic choices of such maps; one represents simultaneous responses and the other represents sequential responses. In what follows, for any subset $a \subset \{1, \dots, J\}$, let θ_{-a} denote the subvector of θ that stacks the components of θ_j 's for all $j \notin a$. If a is a singleton (i.e., $a = \{j\}$ for some j), we simply write θ_{-j} . For each j and $a \subseteq \{1, \dots, J\} \setminus \{j\}$,

let $\pi_{-a} : \Theta_{-j} \rightarrow \prod_{k \in \{1, \dots, J\} \setminus (\{j\} \cup a)} \Theta_k$ be the coordinate projection of θ_{-j} to a further subvector that stacks all components of θ_{-j} except for those of θ_k with $k \in a$.

Let $D_K := \{\theta \in \Theta : \pi_{-j}\theta \in R_{-j}, j = 1, \dots, J\}$. Let $K : D_K \rightarrow \mathbb{R}^{d_X + d_D}$ be a map defined by

$$K(\theta) = \begin{pmatrix} K_1(\theta) \\ \vdots \\ K_J(\theta) \end{pmatrix} = \begin{pmatrix} L_1(\theta_{-1}) \\ \vdots \\ L_J(\theta_{-J}) \end{pmatrix}.$$

This can be interpreted as the players' simultaneous best responses to the initial strategy $(\theta_1, \dots, \theta_J)$. With one endogenous variable, this map simplifies to

$$K(\theta) = \begin{pmatrix} L_1(\theta_2) \\ L_2(\theta_1) \end{pmatrix}.$$

Here, K maps $\theta = (\theta_1, \theta_2)$ to a new parameter value through the simultaneous best responses of players 1 and 2.

Similarly, let $D_M \subset \mathbb{R}^{d_D}$ and let $M : D_M \rightarrow \mathbb{R}^{d_D}$ be a map such that

$$M(\theta_{-1}) = \begin{pmatrix} M_1(\theta_{-1}) \\ M_2(\theta_{-1}) \\ \vdots \\ M_{d_D}(\theta_{-1}) \end{pmatrix} = \begin{pmatrix} L_2(L_1(\theta_{-1}), \theta_{-\{1,2\}}) \\ L_3(L_1(\theta_{-1}), L_2(L_1(\theta_{-1}), \theta_{-\{1,2\}}), \theta_{-\{1,2,3\}}) \\ \vdots \\ L_J(L_1(\theta_{-1}), L_2(L_1(\theta_{-1}), \theta_{-\{1,2\}}), \dots) \end{pmatrix},$$

which can be interpreted as the players' sequential responses (first by player 1, then player 2, etc.) to an initial strategy $\theta_{-1} = (\theta_2, \dots, \theta_J)$.³ Note that the argument of M is not the entire parameter vector. Rather, it is a subvector of θ consisting of the coefficients on the endogenous variables. In order to find a fixed point, this feature is particularly attractive when the number of endogenous variables is small. With one endogenous variable (i.e., $\theta_2 \in \mathbb{R}$ is a scalar), the map simplifies to

$$M(\theta_2) = L_2(L_1(\theta_2)),$$

which is a univariate function whose fixed point is often straightforward to compute.

Define

$$\begin{aligned} \tilde{R}_1 := \{ & \theta_{-1} \in \Theta_{-1} : \Psi_{P,1}(\theta_1, \theta_{-1}) = 0, \\ & \Psi_{P,2}(\theta_1, \theta_2, \pi_{-\{1,2\}}\theta_{-1}) = 0, \exists (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \}. \end{aligned} \tag{3.6}$$

This is the set on which the map $\theta_{-1} \rightarrow L_2(L_1(\theta_{-1}), \pi_{-\{1,2\}}\theta_{-1})$, the first component of M , is well-defined. We then recursively define \tilde{R}_j for $j = 2, \dots, d_D$ in a similar manner.

³One may define M by changing the order of responses as well. For theoretical analysis, it suffices to consider only one of them. Once the fixed point θ_{-1}^* of M is found, one may also obtain θ_1^* using $\theta_1^* = L_1(\theta_{-1}^*)$.

A precise definition of these sets is given in Appendix C. Now define

$$D_M := \bigcap_{j=1}^{d_D} \tilde{R}_j = \tilde{R}_{d_D},$$

where the second equality follows because \tilde{R}_{d_D} turns out to be a subset of \tilde{R}_j for all $j \leq d_D$. The following corollary ensures that K and M are well-defined on D_K and D_M , respectively.

COROLLARY 1. *The maps $K : D_K \rightarrow \mathbb{R}^{d_X+d_D}$ and $M : D_M \rightarrow \mathbb{R}^{d_D}$ exist and are continuously differentiable on the interior of their domains.*

The key insight that we exploit is that, by construction of the BR maps, the problem of finding a solution to $\Psi_P(\theta) = 0$ is equivalent to the problem of finding a fixed point of K (or M). The following proposition states the formal result.

PROPOSITION 1. *Suppose that Assumptions 1 and 2 hold. Then,*

- (i) $\Psi_P(\theta^*) = 0$ if and only if $K(\theta^*) = \theta^*$
- (ii) $\Psi_P(\theta^*) = 0$ if and only if $M(\theta_{-1}^*) = \theta_{-1}^*$ and $\theta_1^* = L_1(\theta_{-1}^*)$.

In view of Proposition 1, the original IVQR estimation problem can be reformulated as the problem of finding the fixed point of K (or M). This reformulation naturally leads to discrete dynamical systems associated with these maps, which in turn provide straightforward iterative algorithms for computing θ^* .

(1) **SIMULTANEOUS DYNAMICAL SYSTEM:**⁴

$$\theta^{(s+1)} = K(\theta^{(s)}), \quad s = 0, 1, 2, \dots, \theta^{(0)} \text{ given.} \tag{3.7}$$

(2) **SEQUENTIAL DYNAMICAL SYSTEM:**⁵

$$\theta_{-1}^{(s+1)} = M(\theta_{-1}^{(s)}), \quad s = 0, 1, 2, \dots, \theta_{-1}^{(0)} \text{ given,} \tag{3.8}$$

where $\theta_1^{(s+1)} = L_1(\theta_{-1}^{(s)})$.

These discrete dynamical systems constitute the basis for our estimation algorithms.⁶

4. POPULATION ALGORITHMS

In this section, we explore the implications of the fixed-point reformulation for constructing population-level algorithms for computing fixed points.

⁴This algorithm is similar to a Jacobi procedure.

⁵Smyth (1996) considers this type of algorithm for $J = 2$ and calls it “zigzag” algorithm. It is similar to a Gauss–Seidel procedure.

⁶These discrete dynamical systems can also be viewed as learning dynamics in a game (e.g., Li and Basar (1987), Fudenberg and Levine (2007)).

4.1 Contraction-based algorithms

We first consider conditions under which K and M are contraction mappings. They ensure that the discrete dynamical systems induced by K and M are convergent to unique fixed points. Moreover, in view of Proposition 1, (point) identification is equivalent to the uniqueness of the fixed point of K or M . Therefore, the conditions we provide below are also sufficient for the point identification of θ^* . We will discuss the relationship between our conditions and existing ones in the next section.

For any vector-valued differentiable map E , let $J_E(x)$ denote its Jacobian matrix evaluated at its argument x . For any matrix A , let $\|A\|$ denote its operator norm. We provide conditions in terms of the Jacobian matrices of K and M , which are well-defined by Corollary 1.

ASSUMPTION 3. *There exist open strictly convex sets $\tilde{D}_K \subseteq D_K$ and $\tilde{D}_M \subseteq D_M$ such that*

- (1) $\|J_K(\theta)\| \leq \lambda$ for some $\lambda < 1$ for all $\theta \in \tilde{D}_K$;
- (2) $\|J_M(\theta_{-1})\| \leq \lambda$ for some $\lambda < 1$ for all $\theta_{-1} \in \tilde{D}_M$.

Under this additional assumption, the iterative algorithms are guaranteed to converge to the fixed point. We summarize this result below.

PROPOSITION 2. *Suppose that Assumptions 1 and 2 hold.*

- (i) *Suppose further that Assumption 3(1) holds. Then K is a contraction on the closure of \tilde{D}_K . The fixed point $\theta^* \in \text{cl}(\tilde{D}_K)$ of K is unique. For any $\theta^{(0)} \in \tilde{D}_K$, the sequence $\{\theta^{(s)}\}_{s=0}^\infty$ defined in (3.7) satisfies $\theta^{(s)} \rightarrow \theta^*$ as $s \rightarrow \infty$.*
- (ii) *Suppose further that Assumption 3(2) holds. Then M is a contraction on the closure of \tilde{D}_M . The fixed point $\theta_{-1}^* \in \text{cl}(\tilde{D}_M)$ of M is unique. For any $\theta_{-1}^{(0)} \in \tilde{D}_M$, the sequence $\{\theta_{-1}^{(s)}\}_{s=0}^\infty$ defined in (3.8) satisfies $\theta_{-1}^{(s)} \rightarrow \theta_{-1}^*$ as $s \rightarrow \infty$.*

In the case of a single endogenous variable, the Jacobian matrices of K and M are given by

$$J_K(\theta) = \begin{pmatrix} 0 & J_{L_1}(\theta_2) \\ J_{L_2}(\theta_1) & 0 \end{pmatrix} \quad \text{and} \quad J_M(\theta_2) = J_{L_2}(L_1(\theta_2))J_{L_1}(\theta_2),$$

where

$$J_{L_{-j}}(\theta_j) = - \left(\frac{\partial \Psi_{P,-j}(\theta_j, \theta_{-j})}{\partial \theta'_{-j}} \Bigg|_{\theta=(\theta_j, L_{-j}(\theta_j))} \right)^{-1} \times \frac{\partial \Psi_{P,-j}(\theta_j, \theta_{-j})}{\partial \theta'_j} \Bigg|_{\theta=(\theta_j, L_{-j}(\theta_j))}, \quad j = 1, 2. \tag{4.1}$$

One may therefore check the high-level condition through the Jacobians of the original moment restrictions. In Appendix C.2.1, we illustrate a simple primitive condition for

a local version of Assumption 3. In practice, we found that violations of Assumption 3 lead to explosive behavior of the estimation algorithms, and thus, are very easy to detect numerically.

4.2 Connections to the identification conditions in the literature

In view of Proposition 1, identification of θ^* is equivalent to uniqueness of the fixed points of K and M , which is ensured by Proposition 2. Here, we discuss how the conditions required by Proposition 2 relate to the ones in the literature.

We start with local identification. The parameter vector θ^* is said to be locally identified if there is a neighborhood \mathcal{N} of θ^* such that $\Psi_P(\theta) \neq 0$ for all $\theta \neq \theta^*$ in the neighborhood. Local identification in the IVQR model follows from standard results (e.g., [Rothenberg \(1971\)](#), [Chen, Chernozhukov, Lee, and Newey \(2014\)](#)). For example, if $\Psi_P(\theta)$ is differentiable, [Chen et al. \(2014, Section 2.1\)](#) show that full rank of $J_{\Psi_P}(\theta)$ at θ^* is sufficient for local identification.

It is interesting to compare this full rank condition to Assumption 5(1) in the Appendix, which is a local version of Assumption 3(1). Assumption 5(1) requires that $\rho(J_K(\theta^*)) < 1$, where $\rho(A)$ denotes the spectral radius of a square matrix A . We highlight the connection in the case with a single endogenous variable. Full rank of $J_{\Psi_P}(\theta^*)$ is equivalent to $\det(J_{\Psi_P}(\theta^*)) \neq 0$. Observe that, for any θ ,

$$\begin{aligned} \det(J_{\Psi_P}(\theta)) &= \det \begin{pmatrix} \partial\Psi_{P,1}(\theta_1, \theta_2)/\partial\theta'_1 & \partial\Psi_{P,1}(\theta_1, \theta_2)/\partial\theta'_2 \\ \partial\Psi_{P,2}(\theta_1, \theta_2)/\partial\theta'_1 & \partial\Psi_{P,2}(\theta_1, \theta_2)/\partial\theta'_2 \end{pmatrix} \\ &= \det \left(\begin{pmatrix} \partial\Psi_{P,1}(\theta_1, \theta_2)/\partial\theta'_1 & 0 \\ 0 & \partial\Psi_{P,2}(\theta_1, \theta_2)/\partial\theta'_2 \end{pmatrix} \begin{pmatrix} I_{d_1} & -J_{L_1}(\theta_2) \\ -J_{L_2}(\theta_1) & I_{d_2} \end{pmatrix} \right) \\ &= \det \begin{pmatrix} \partial\Psi_{P,1}(\theta_1, \theta_2)/\partial\theta'_1 & 0 \\ 0 & \partial\Psi_{P,2}(\theta_1, \theta_2)/\partial\theta'_2 \end{pmatrix} \det \begin{pmatrix} I_{d_1} & -J_{L_1}(\theta_2) \\ -J_{L_2}(\theta_1) & I_{d_2} \end{pmatrix}. \end{aligned}$$

If $\partial\Psi_{P,j}(\theta)/\partial\theta'_j|_{\theta=\theta^*}$ is invertible for $j = 1, 2$ (which is true under Assumption 2(4)), $J_{\Psi_P}(\theta^*)$ is full rank if and only if

$$0 \neq \det \begin{pmatrix} I_{d_1} & -J_{L_1}(\theta_2^*) \\ -J_{L_2}(\theta_1^*) & I_{d_2} \end{pmatrix} = \det(I_{d_X+d_D} - J_K(\theta^*)). \tag{4.2}$$

That is, it requires that none of the eigenvalues of J_K has modulus one. Therefore, Assumption 5(1) is sufficient but not necessary for condition (4.2) to hold. Specifically, Assumption 5(1) requires all eigenvalues of $J_K(\theta^*)$ to lie strictly within the unit circle, while local identification only requires all eigenvalues not to be on the unit circle. In terms of the dynamical system induced by K , the former ensures that the dynamical system has a unique *asymptotically stable fixed point*, while the latter ensures that the system has a unique *hyperbolic fixed point*, which is a more general class of fixed points (e.g., [Galor \(2007\)](#)).⁷ Under the former condition, iteratively applying the contraction map induces

⁷The argument above also applies to settings with multiple endogenous variables. A similar result can also be shown for M .

convergence, while the latter generally requires a root finding method to obtain the fixed point.

Now we turn to global identification and compare Proposition 2 to the global identification result in Chernozhukov and Hansen (2006).

LEMMA 2 (Theorem 2 in Chernozhukov and Hansen (2006)). *Suppose that Assumption 1 holds. Moreover, suppose that (i) Θ is compact and convex and θ^* is in the interior of Θ ; (ii) $f_{Y|D,Z,X}$ is uniformly bounded a.s.; (iii) $J_{\Psi_P}(\theta)$ is continuous and has full rank uniformly over Θ ; and (iv) the image of Θ under the mapping $\theta \rightarrow \Psi_P(\theta)$ is simply connected. Then θ^* uniquely solves $\Psi_P(\theta) = 0$ over Θ .*

Under Conditions (i)–(iv), which are substantially stronger than the local identification conditions discussed above, the result in Lemma 2 follows from an application of Hadamard’s global univalence theorem (e.g., Theorem 1.8 in Ambrosetti and Prodi (1995)).

Comparing Lemma 2 to Proposition 2, we can see that the result in Lemma 2 establishes identification over the whole parameter space Θ , while Proposition 2 establishes identification over the sets \tilde{D}_K and \tilde{D}_M , which will generally be subsets of Θ . Regarding the underlying assumptions, Conditions (i) and (ii) in Lemma 2 correspond to our Assumptions 2(1) and 2(3). Moreover, our Assumption 2(3) constitutes an easy-to-interpret sufficient condition for continuity of J_{Ψ_P} as required in Condition (iii). To apply Hadamard’s global univalence theorem, Chernozhukov and Hansen (2006) assume the simple connectedness of the image of Ψ_P (Condition (iv)). By contrast, we use a different univalence theorem by Gale and Nikaido (1965) (applied to the map Ξ defined in (E.1) that arises from each subsystem), which does not require further conditions. However, when establishing global identification based on the contraction mapping theorem, we need to impose an additional condition on the Jacobian (Assumption 3). In sum, our conditions are somewhat stronger in terms of restrictions on the Jacobian, but they are relatively easy to check and allow us to dispense with an abstract condition (simple connectedness of the image of a certain map) to apply a global univalence theorem.

4.3 Root-finding algorithms and nesting

Assumption 3 is a sufficient condition for the uniqueness of the fixed point and the convergence of the contraction-based algorithms. Even in settings where this assumption fails to hold, one may still identify θ^* and design an algorithm that is able to find it under weaker conditions on the Jacobian. This is the case under the assumptions in the general (global) identification result of Chernozhukov and Hansen (2006); see Lemma 2.

Note that, for the simultaneous dynamical system, θ^* solves

$$(I_{d_X+d_D} - K)(\theta^*) = 0,$$

where $I_{d_X+d_D}$ is the identity map. Similarly, in the sequential dynamical system, θ_{-1}^* solves

$$(I_{d_D} - M)(\theta_{-1}^*) = 0.$$

Therefore, standard root-finding algorithms can be used to compute the fixed point.

For implementing root-finding algorithms, we find that reducing the dimension of the fixed-point problem is often helpful. Toward this end, we briefly discuss another class of dynamical systems and associated population algorithms which can be used for the purpose of dimension reduction. Namely, with more than two players, one can construct *nested* dynamical systems, which induce nested fixed-point algorithms. Nesting is useful as it allows for transforming any setup with more than two players into a two-player system.

To fix ideas, consider the case of three players ($J = 3$). Fix player 3's action $\theta_3 \in \Theta_3 \subset \mathbb{R}$ and consider the associated “subgame” between players 1 and 2. To describe the subgame, define $M_{1,2|3}(\cdot | \theta_3) : \Theta_2 \rightarrow \Theta_2$ pointwise by

$$M_{1,2|3}(\theta_2 | \theta_3) := L_2(L_1(\theta_2, \theta_3), \theta_3).$$

This map gives the sequential best responses of players 1 and 2, while taking player 3's strategy given. Define the fixed point $L_{12} : \Theta_3 \rightarrow \Theta_1 \times \Theta_2$ of the subgame by

$$L_{12}(\theta_3) := \begin{pmatrix} \bar{\theta}_1(\theta_3) \\ \bar{\theta}_2(\theta_3) \end{pmatrix} = \begin{pmatrix} L_1(\bar{\theta}_2(\theta_3), \theta_3) \\ M_{1,2|3}(\bar{\theta}_2(\theta_3) | \theta_3) \end{pmatrix}.$$

This map then defines a new “best response” map. Here, given θ_3 , the players in the subgame (i.e., players 1 and 2) collectively respond by choosing the Nash equilibrium of the subgame. The overall dynamical system induced by the nested decentralization is then given by

$$M_3(\theta_3) = L_3(L_{12}(\theta_3)).$$

Hence, we can interpret the nested algorithm as a two-player dynamical system where one player solves an internal fixed-point problem. The nesting algorithms require existence and uniqueness of the fixed points in the subgame between players 1 and 2. In Appendix C.3, we discuss the formal conditions required for the existence and uniqueness of such fixed points.⁸

This nesting procedure is generic and can be extended to more than three players by sequentially adding additional layers of nesting.⁹ It follows that any decentralized estimation problem with more than two players can be reformulated as a nested dynamical system with two players: player J and all others $-J$. The resulting dynamical system $M_J(\theta_J) = L_J(L_{-J}(\theta_J))$ is particularly useful when M_J is not necessarily a contraction map since θ_J is a scalar such that, as we see below, its fixed point can be computed using univariate root-finding algorithms.

⁸For an equilibrium of the subgame to be well-defined, one may directly assume that Chernozhukov and Hansen's (2006) global identification condition holds within the subgame, given player 3's action θ_3 . Alternatively, if Assumption 3 holds, it implies that $M_{1,2|3}(\cdot | \theta_3)$ is a contraction, which in turn ensures the existence and uniqueness of the fixed points in the subgame; see Appendix C.3.

⁹In the current example, consider adding player 4 and letting players 1–3 best respond by returning the fixed point of the subgame through M_3 given θ_4 . One can repeat this for additional players. This procedure can also be applied to the simultaneous dynamical system induced by K .

5. SAMPLE ESTIMATION ALGORITHMS

5.1 Sample estimation problem

Let $\{(Y_i, D'_i, X'_i, Z'_i)\}_{i=1}^N$ be a sample generated from the IVQR model. Our estimators are constructed using the analogy principle. For this, define the sample payoff functions for the players as

$$Q_{N,1}(\theta) := \frac{1}{N} \sum_{i=1}^N \rho_\tau(Y_i - X'_i\theta_1 - D_{1,i}\theta_2 - \dots - D_{d_D,i}\theta_J), \quad (5.1)$$

$$Q_{N,j}(\theta) := \frac{1}{N} \sum_{i=1}^N \rho_\tau(Y_i - X'_i\theta_1 - D_{1,i}\theta_2 - \dots - D_{d_D,i}\theta_J) \frac{Z_{j-1,i}}{D_{j-1,i}}, \quad j = 2, \dots, J. \quad (5.2)$$

For each $j = 1, \dots, J$, let the sample BR function $\hat{L}_j(\theta_{-j})$ be a function such that

$$\hat{L}_1(\theta_{-1}) \in \arg \min_{\theta_1 \in \mathbb{R}^{d_X}} Q_{N,1}(\tilde{\theta}_1, \theta_{-1}), \quad (5.3)$$

$$\hat{L}_j(\theta_{-j}) \in \arg \min_{\theta_j \in \mathbb{R}} Q_{N,j}(\tilde{\theta}_j, \theta_{-j}), \quad j = 2, \dots, J. \quad (5.4)$$

Assuming that the model is parametrized in such a way that $Z_{\ell,i}/D_{\ell,i}$, $\ell = 1, \dots, d_D$, is positive, these are convex (weighted) QR problems for which fast solution algorithms exist. In our empirical applications and simulations, we use the R-package `quantreg` to estimate the QRs (Koenker (2020)). For example, $\hat{L}_2(\theta_{-2})$ can be computed by running a QR with weights $Z_{1,i}/D_{1,i}$ in which one regresses $Y_i - X'_i\theta_1 - D_{2,i}\theta_3 - \dots - D_{d_D,i}\theta_J$ on $D_{1,i}$ without a constant. These sample BR functions also approximately solve the sample analog of the moment restrictions in (3.3)–(3.4); see Lemma 10 in the Appendix.

REMARK 5.1. The proposed estimators rely on decentralizing the original nonsmooth and nonconvex IVQR GMM problem into a series of convex QR problems. The quality and the computational performance of our procedures therefore crucially depends on the choice of the underlying QR estimation approach, which deserves some further discussion. The interested reader is referred to Koenker (2017) for an excellent overview on the computational aspects of QR. In this paper, we use the (modified) Barrodale and Roberts algorithm, which is implemented as the default in the `quantreg` package and described in Koenker and D'Orey (1987, 1994). The `quantreg` implementation of this algorithm is computationally tractable for problems up to several thousand observations. For larger problems, we recommend using interior point methods, potentially after pre-processing; see Portnoy and Koenker (1997) for a detailed description. These methods are conveniently implemented in the `quantreg` package. For very large problems, one can resort to first-order gradient descent methods, which are amenable to modern parallelized computation; see Section 5.5 in Koenker (2017) for an introduction and simulation evidence on the performance of such methods.

We focus on estimators constructed based on the dynamical system M . Lemma 5 in the Appendix shows that the estimators based on M and K are asymptotically equivalent. However, in our simulations, we found that, while the convergence properties of contraction algorithms based on M are excellent, the convergence properties of contraction algorithms based on K can be sensitive to the choice of starting value. This may be attributed to the fact that the algorithm based on K requires all $d_X + d_D$ components of the starting value to be in the domain of the contraction map, while the algorithm based on M only requires that the same condition is satisfied by the starting value for θ_{-1} . Our simulations suggest that it is often easier to satisfy this requirement with M since the number of components in θ_{-1} , d_D , is typically much smaller than $d_X + d_D$. Also, for root-finding algorithms, we prefer the sequential dynamical system (induced by M) because it again leads to a substantial dimension reduction: it reduces the original $(d_X + d_D)$ -dimensional GMM estimation problem to a d_D -dimensional root-finding problem.

We construct estimation algorithms by mimicking the population algorithms. Let \hat{M} denote a sample analog of M :

$$\hat{M}(\theta_{-1}) := \begin{pmatrix} \hat{M}_1(\theta_{-1}) \\ \hat{M}_2(\theta_{-1}) \\ \vdots \\ \hat{M}_{d_D}(\theta_{-1}) \end{pmatrix} = \begin{pmatrix} \hat{L}_2(\hat{L}_1(\theta_{-1}), \theta_{-(1,2)}) \\ \hat{L}_3(\hat{L}_1(\theta_{-1}), \hat{L}_2(\hat{L}_1(\theta_{-1}), \theta_{-(1,2)}), \theta_{-(1,2,3)}) \\ \vdots \\ \hat{L}_J(\hat{L}_1(\theta_{-1}), \hat{L}_2(\hat{L}_1(\theta_{-1}), \theta_{-(1,2)}), \dots) \end{pmatrix},$$

where $\theta_1 = \hat{L}_1(\theta_{-1})$. This map induces a sample analog of the sequential dynamical system in Section 3.

SAMPLE SEQUENTIAL DYNAMICAL SYSTEM:

$$\theta_{-1}^{(s+1)} = \hat{M}(\theta_{-1}^{(s)}), \quad s = 0, 1, 2, \dots, \theta_{-1}^{(0)} \text{ given}, \tag{5.5}$$

where $\theta_1^{(s+1)} = \hat{L}_1(\theta_{-1}^{(s)})$.

5.2 Contraction-based algorithms

The first set of algorithms exploits that, under Assumption 3, \hat{M} is a contraction mapping with probability approaching one. In this case, we iterate the dynamical system (5.5) until $\|\theta_{-1}^{(s)} - \hat{M}(\theta_{-1}^{(s)})\|$ is within a numerical tolerance e_N .¹⁰ This iterative algorithm is known to converge at least linearly. The approximate sample fixed point $\hat{\theta}_N = (\hat{\theta}_{N,1}, \hat{\theta}_{N,-1})$ that meets the convergence criterion then serves as an estimator for θ .

5.3 Algorithms based on root-finders and optimizers

We construct an estimator $\hat{\theta}_N$ of θ^* as an approximate fixed point to the sample problem:

$$\|\hat{\theta}_{N,-1} - \hat{M}(\hat{\theta}_{N,-1})\| \leq e_N,$$

¹⁰In the next section, we require $e_N = o(N^{-1/2})$, which ensures that the numerical error does not affect the asymptotic distribution.

where $\hat{\theta}_{N,1} = \hat{L}_1(\hat{\theta}_{N,-1})$ and e_N is a numerical tolerance. This problem can be solved efficiently using well-established root-finding algorithms since \hat{M} is easy to evaluate as the composition of standard QRs. When $d_D = 1$, one may use Brent's method (Brent (1971)) whose convergence is superlinear. When $d_D > 1$, one could apply the Newton–Raphson method, which achieves quadratic convergence but requires an estimate or a finite difference approximation of the derivative. The corresponding approximation error may affect the performance. Alternatively, one can compute the fixed point by minimizing $\|\hat{M}(\theta) - \theta\|^2$. The potential issue with this approach is that translating the root-finding problem into a minimization problem can lead to local minima in the objective function. Therefore, it is important to use global optimization strategies.

As described in Section 4.3, nesting can be used to reduce the dimensionality even further. In particular, the problem can be reformulated as a one-dimensional fixed-point problem, which can be solved using existing methods. We found that Brent's method performs very well in our context. Nesting is suitable when the number of endogenous variables is moderate. While adding nests incurs additional computational costs, our Monte Carlo experiments suggest that they are not excessive when the number of endogenous variables is moderate.¹¹

5.4 Profiling algorithms

A key insight underlying our estimation algorithms is that, given θ_{-1} , the estimation problem becomes a standard convex QR problem:

$$\hat{L}_1(\theta_{-1}) \in \arg \min_{\tilde{\theta}_1 \in \mathbb{R}^{d_X}} Q_{N,1}(\tilde{\theta}_1, \theta_{-1}).$$

This insight suggests a profiling estimator.¹² In particular, $\hat{\theta}_{N,-1}$ can be obtained as the approximate root of the function

$$f(\theta_{-1}) = \frac{1}{N} \sum_{i=1}^N (1\{Y_i \leq X_i' \hat{L}_1(\theta_{-1}) + D_i' \theta_{-1}\} - \tau) Z_i,$$

and $\hat{\theta}_{N,1}$ can be estimated as $\hat{\theta}_{N,1} \in \arg \min_{\tilde{\theta}_1 \in \mathbb{R}^{d_X}} Q_{N,1}(\tilde{\theta}_1, \hat{\theta}_{N,-1})$. When there is only one endogenous variable, f is scalar-valued and univariate root-finders can be used. With multivariate endogenous variables, one can either directly apply multivariate root-finders or construct nested algorithms as described in Sections 4.3 and 5.3.

Relative to the root-finding methods described in Section 5.3, the profiling estimator has the advantage that evaluating f only requires estimating one QR. On the other hand, the root-finding algorithms in Section 5.3 efficiently exploit the convexity of the subproblems for players $j = 2, \dots, J$, and demonstrate a better computational performance than the profiling estimators with multiple endogenous variables (cf. Table 3).

¹¹See Section 8 and Appendix D.

¹²We thank an anonymous referee for suggesting a comparison to this alternative estimator. Chernozhukov, Hansen, and Wüthrich (2017) use profiling to construct IVQR GMM estimators based on Neyman orthogonal scores.

6. ASYMPTOTIC THEORY

6.1 Estimators

We define an estimator $\hat{\theta}_N$ of θ^* as an approximate fixed point of \hat{M} in the following sense:

$$\|\hat{\theta}_{N,-1} - \hat{M}(\hat{\theta}_{N,-1})\| \leq \inf_{\theta'_{-1} \in \Theta_{-1}} \|\theta'_{-1} - \hat{M}(\theta'_{-1})\| + o_p(N^{-1/2}). \tag{6.1}$$

An estimator of θ_1^* can be constructed by setting

$$\hat{\theta}_{N,1} := \hat{L}_1(\hat{\theta}_{N,-1}). \tag{6.2}$$

In what follows, we call $\hat{\theta}_N = (\hat{\theta}_{N,1}, \hat{\theta}_{N,-1})$ the *fixed-point estimator* of θ^* . Under the conditions we introduce below, $\hat{\theta}_N$ is also an approximate fixed point of \hat{K} ; see Lemma 5 in the Appendix. This turns out to be useful for stating theoretical results in a concise manner. While we mostly focus on algorithms based on \hat{M} below, some of our theoretical results will be stated using K . \hat{M} (or \hat{K}) is defined similarly for the nested dynamical system in which one player solves a fixed-point problem in a subgame.

Consistency and parametric convergence rates of $\hat{\theta}_N$ can be established using existing results. When \hat{M} is asymptotically a contraction map, one may construct an estimator $\hat{\theta}_N$ satisfying (6.1)–(6.2) using the contraction algorithm in Section 5.2 with tolerance $e_N = o(N^{-1/2})$. One may then apply the result of [Dornik and Sherman \(2005\)](#) to obtain the root- N consistency of the estimator.¹³ For completeness, this result is summarized in Appendix H.

More generally, if \hat{M} is not guaranteed to be a contraction, one may use root-finding algorithms that solve $\theta_{-1} - \hat{M}(\theta_{-1}) = 0$ up to approximation errors of $o(N^{-1/2})$. The root- N consistency of $\theta_{N,-1}$ then follows from the standard argument for extremum estimators, in which we take $Q_N(\theta_{-1}) = \|\theta_{-1} - \hat{M}(\theta_{-1})\|$ as a criterion function.¹⁴ Since these results are standard, we omit details and focus below on the asymptotic distribution and bootstrap validity of the fixed-point estimators. Our contributions are two-fold. First, we establish the asymptotic distribution of the fixed-point estimator without assuming that \hat{M} is an asymptotic contraction map, which therefore allows the practitioner to conduct inference using the estimator based on the general root-finding algorithm and complements the result of [Dornik and Sherman \(2005\)](#). Second, to our knowledge, the bootstrap validity of the fixed-point estimators is new. These results are established by showing that, under regularity conditions, the population fixed point is Hadamard-differentiable, and hence admits the use of the functional δ -method, which may be of independent theoretical interest.

REMARK 6.1. To establish the asymptotic properties, one could try to reformulate our estimator as an estimator that approximately solves a GMM problem. Here, instead of

¹³Satisfying $e_N = o(N^{-1/2})$ requires the number of iterations to increase as the sample size tends to infinity, which in turn satisfies requirement (ii) in Theorem 2 of [Dornik and Sherman \(2005\)](#).

¹⁴The key conditions for these results, uniform convergence (in probability) of the sample BR maps and their stochastic equicontinuity, are established in Lemma 11.

relying on another reformulation, which would require establishing a sample analog version of Proposition 1, we develop and directly apply an asymptotic theory for fixed-point estimators. We take this approach as the theory itself contains generic results (Theorem 1 and Lemmas 6–7) surrounding the Hadamard-differentiability of fixed points, which allow for applying the functional δ -method to obtain the asymptotic distribution of $\hat{\theta}_N$ and bootstrap validity. These results can potentially be used to analyze decentralized estimators outside the IVQR class.

6.2 Asymptotic theory and bootstrap validity

The following theorem gives the limiting distribution of our estimator. For each $w = (y, d', x', z)'$ and $\theta \in \Theta$, let $f(w; \theta) \in \mathbb{R}^{d_X+d_D}$ be a vector whose subvectors are given by

$$\begin{aligned} f_1(w; \theta) &= (1\{y \leq d'\theta_{-1} + x'\theta_1\} - \tau)x, \\ f_j(w; \theta) &= (1\{y \leq d'\theta_{-1} + x'\theta_1\} - \tau)z_{j-1}, \quad j = 2, \dots, J, \end{aligned}$$

and let $g(w; \theta) = (g_1(w; \theta)', \dots, g_J(w; \theta)')$ be a vector such that

$$g_j(w; \theta) = \left(\frac{\partial^2}{\partial \theta_j \partial \theta_j'} Q_{P,j}(L_j(\theta_{-j}), \theta_{-j}) \right)^{-1} f_j(w; L_j(\theta_{-j}), \theta_{-j}), \quad j = 1, \dots, J.$$

THEOREM 1. *Suppose that Assumptions 1 and 2 hold. Suppose further that $J_{\Psi_P}(\theta^*)$ is full rank. Let $\{W_i\}_{i=1}^N$ be an i.i.d. sample generated from the IVQR model, where $W_i = (Y_i, D'_i, X'_i, Z'_i)$. Then*

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \xrightarrow{L} N(0, V), \tag{6.3}$$

with

$$V = (I_{d_X+d_D} - J_K(\theta^*))^{-1} E[\mathbb{W}(\theta^*)\mathbb{W}(\theta^*)'] [(I_{d_X+d_D} - J_K(\theta^*))^{-1}]', \tag{6.4}$$

where \mathbb{W} is a tight Gaussian process in $\ell^\infty(\Theta)^{d_X+d_D}$ with mean zero and the covariance kernel

$$\text{Cov}(\mathbb{W}(\theta), \mathbb{W}(\tilde{\theta})) = E_P[(g(W; \theta) - E_P[g(W; \theta)])(g(W; \tilde{\theta}) - E_P[g(W; \tilde{\theta})])'].$$

In Theorem 1, we assume $J_{\Psi_P}(\theta^*)$ is full rank, which ensures local identification of θ^* . The asymptotic variance of $\hat{\theta}_N$ is characterized in terms of the Jacobian of K and a Gaussian process \mathbb{W} . This can be reformulated to show its asymptotic relationship to a GMM estimator (which is asymptotically equivalent to IQR). Let $\tilde{\theta}_N$ be an estimator that solves the following estimating equations:

$$\frac{1}{N} \sum_{i=1}^N 1\{Y_i \leq X'_i \tilde{\theta}_{N,1} + D_{1,i} \tilde{\theta}_{N,2} + \dots + D_{d_D,i} \tilde{\theta}_{N,J}\} - \tau \begin{pmatrix} X_i \\ Z_i \end{pmatrix} = o_p(N^{-1/2}). \tag{6.5}$$

Let $\Psi(\tau) = (X', Z)'$. As shown in Chernozhukov and Hansen (2006, Section 3.3, Theorem 3, and Remark 3), $\sqrt{N}(\hat{\theta}_N - \theta^*)$ converges weakly to a mean zero multivariate normal distribution with variance

$$\tilde{V} = \tau(1 - \tau)J_{\Psi_p}(\theta^*)^{-1}E[\Psi(\tau)\Psi(\tau)'] [J_{\Psi_p}(\theta^*)^{-1}]', \tag{6.6}$$

where $J_{\Psi_p}(\theta^*) = E[f_{\varepsilon(\tau)|X,D,Z}(0)\Psi(\tau)(X', D)']$, $\varepsilon(\tau) = Y - X'\theta_1^* - D'\theta_{-1}^*$, and $f_{\varepsilon(\tau)|X,D,Z}$ is $\varepsilon(\tau)$'s conditional density given (X, D, Z) . The following corollary shows that the fixed-point estimator $\hat{\theta}_N$ is asymptotically equivalent to $\tilde{\theta}_N$ in terms of its limiting distribution.

COROLLARY 2. *Suppose the conditions of Theorem 1 hold. Then $V = \tilde{V}$.*

To conduct inference on θ^* , one may employ a natural bootstrap procedure. For this, use in (6.1)–(6.2) the bootstrap sample instead of the original sample to define the bootstrap analogs \hat{M}^* and $\hat{\theta}_{N,-1}^*$ of \hat{M} and $\hat{\theta}_{N,-1}$. In practice, the bootstrap can be implemented using the following steps:

- (1) Compute the fixed-point estimator $\hat{\theta}_N$ using the original sample.
- (2) Draw a bootstrap sample $\{W_i^*\}_{i=1}^N$ randomly with replacement from the empirical distribution P_N . Use the sequential dynamical system based on \hat{M}^* combined with a contraction or root-finding algorithm to compute $\hat{\theta}_N^*$.
- (3) Repeat Step 2 across bootstrap replications $b = 1, \dots, B$. Let

$$F_B(x) := \frac{1}{B} \sum_{b=1}^B 1\{\sqrt{N}(\hat{\theta}_N^{*,b} - \hat{\theta}_N) \leq x\}, \quad x \in \mathbb{R}.$$

Use F_B as an approximation to the sampling distribution of the root $\sqrt{N}(\hat{\theta}_N - \theta^*)$.

The bootstrap is particularly attractive in conjunction with our new and computationally efficient estimation algorithms. By contrast, directly bootstrapping for instance the IQR estimator of Chernozhukov and Hansen (2006) is computationally very costly. Alternative methods (either an asymptotic approximation or a score-based bootstrap) require estimation of the influence function, which involves nonparametric estimation of a certain conditional density. Directly bootstrapping our fixed-point estimators avoids the use of any smoothing and tuning parameters.¹⁵

The following theorem establishes the consistency of the bootstrap procedure. For this, let $\overset{L^*}{\rightsquigarrow}$ denote the weak convergence of the bootstrap law in outer probability, conditional on the sample path $\{W_i\}_{i=1}^\infty$.

¹⁵The use of the bootstrap here is for consistently estimating the law of the estimator. Whether one may obtain higher-order refinements through a version of the bootstrap, for example, the m out of n bootstrap with extrapolation (Sakov and Bickel (2000)), is an interesting question which we leave for future research.

THEOREM 2. *Suppose that Assumptions 1 and 2 hold. Suppose further that $J_{\Psi_P}(\theta^*)$ is full rank. Let $\{W_i\}_{i=1}^N$ be an i.i.d. sample generated from the IVQR model. Then*

$$\sqrt{N}(\hat{\theta}_N^* - \hat{\theta}_N) \overset{L^*}{\rightsquigarrow} N(0, V),$$

where V is as in (6.4).

7. EMPIRICAL EXAMPLE

In this section, we illustrate the proposed estimators by reanalyzing the effect of 401(k) plans on savings behavior as in Chernozhukov and Hansen (2004). This empirical example constitutes the basis for our Monte Carlo simulations in Section 8. As described by Chernozhukov and Hansen (2004), 401(k) plans are tax-deferred savings options that allow for deducting contributions from taxable income and accruing tax-free interest. These plans are provided by employers and were introduced in the United States in the early 1980s to increase individual savings. To estimate the effect of 401(k) plans (D) on accumulated assets (Y), one has to deal with the potential endogeneity of 401(k) participation. Chernozhukov and Hansen (2004) proposed an instrumental variables approach to overcome this problem. They use 401(k) eligibility as an instrument (Z) for the participation in 401(k) plans. The underlying argument, which is due to Poterba, Venti, and Wise (1994, 1995, 1998) and Benjamin (2003), is that eligibility is exogenous after conditioning on income and other observable covariates. We use the same empirical approach here but note that there are also papers arguing that 401(k) eligibility violates conditional exogeneity (e.g., Engen, Gale, and Scholz (1996)).

We use the same dataset as in Chernozhukov and Hansen (2004). The dataset contains information about 9913 observations from a sample of households from the 1991 Survey of Income and Program Participation.¹⁶ We refer to Chernozhukov and Hansen (2004) for more information about the data and to their Tables 1 and 2 for descriptive statistics. Here, we focus on net financial assets as our outcome of interest.¹⁷

Following Chernozhukov and Hansen (2004), we consider the following linear IVQR model:

$$q(D, X, \tau) = X' \theta_X(\tau) + D \theta_D(\tau).$$

The vector of covariates X includes seven dummies for income categories, five dummies for age categories, family size, four dummies for education categories, indicators for marital status, two-earner status, defined benefit pension status, individual retirement account participation status and homeownership, and a constant. Because $P(D = 0) > 0$, we re-parametrize the model by replacing D by $D^* = D + 1$ to ensure that Z/D^* is well-defined and positive.

Below, we briefly describe the construction of the sequential response map for this application. First, we allocate θ_X to player 1, and hence denote this subvector by θ_1 . Similarly, we allocate θ_D to player 2 and denote it by θ_2 . For each τ , the sequential response map can be constructed by taking the following steps:

¹⁶The dataset analyzed by Chernozhukov and Hansen (2004) has 9915 observations. Here, we delete the two observations with negative income.

¹⁷Chernozhukov and Hansen (2004) also considered total wealth and net nonfinancial assets.

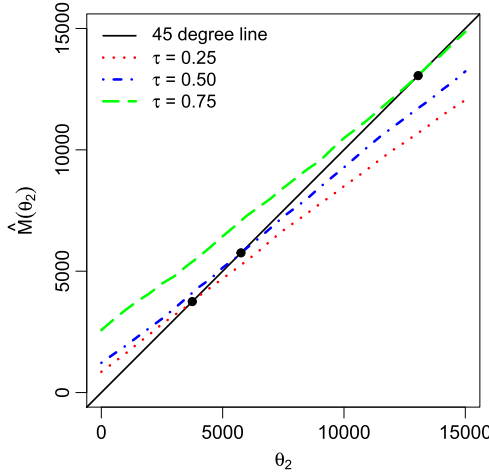


FIGURE 1. Illustration fixed-point estimator.

- (1) For any given θ_2 , compute $\theta_1 = \hat{L}_1(\theta_2)$ by running an ordinary QR (for the τ th quantile) using $Y_i - D_i^* \theta_2$ as the dependent variable and X_i as a regressor vector.
- (2) Given $\theta_1 = \hat{L}_1(\theta_2)$ from the previous step, compute $\hat{L}_2(\theta_1)$ by running an ordinary QR with weights Z_i/D_i^* using $Y_i - X_i' \theta_1$ as the dependent variable, D_i^* as the regressor, and omitting the constant.

Combining these two steps yields the sequential response map $\hat{M}(\cdot) = \hat{L}_2(\hat{L}_1(\cdot))$. Figure 1 shows, for each $\tau \in \{0.25, 0.50, 0.75\}$, the graph of $\theta_2 \mapsto \hat{M}(\theta_2)$. For each τ , the intersection between \hat{M} and the 45-degree line (i.e., the identity map) is our fixed-point estimator $\hat{\theta}_D(\tau)$. Figure 1 further provides a straightforward graphical way to check the validity of the sample analog of Assumption 3. We can see that the sample analog of J_M (i.e., the slope of M) is smaller than one at any θ_2 . This suggests that the contraction-based sequential algorithm converges at all three quantile levels, which is indeed what we find.

For each τ , the steps for numerically calculating the fixed-point estimator are as follows:

- (3) Find the fixed point of \hat{M} using one of the following algorithms:
 - Contraction algorithm: Set the tolerance e_N .¹⁸ We set the initial value $\theta_2^{(0)}$ to the 2SLS estimate of the coefficient on D_i^* . Iterate $\theta_2^{(s+1)} = \hat{M}(\theta_2^{(s)})$, for $s = 0, 1, 2, \dots$, until $|\theta_2^{(s+1)} - \theta_2^{(s)}| \leq e_N$. Report $\theta_2^{(s+1)}$ as the estimate $\hat{\theta}_D$ of θ_D .
 - Root-finding algorithm: Find the solution to the equation $\theta_2 - \hat{M}(\theta_2) = 0$ by applying Brent's method as implemented by the R-function `uniroot`. Report the solution as the estimate $\hat{\theta}_D$ of θ_D .

¹⁸In our application and simulations, we set e_N equal to the square-root of the machine precision in R: $e_N \approx 1.5 \cdot 10^{-8}$.

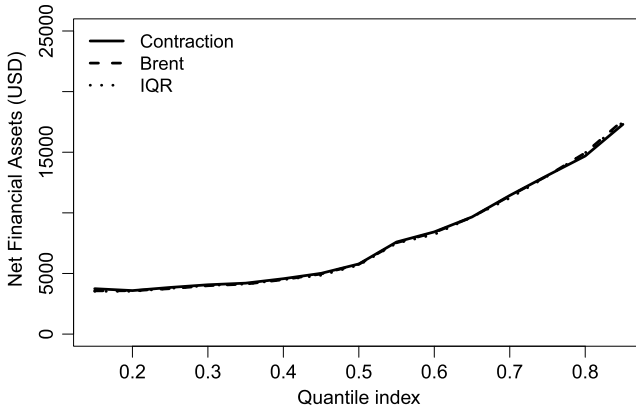


FIGURE 2. Comparison point estimates.

- (4) (Optional) Compute $\hat{\theta}_X = \hat{L}_1(\hat{\theta}_D)$ (as in Step (2)) if one is also interested in the coefficients on X . Recover estimates of the original coefficients following the discussion in Appendix B.

We compare our estimators to the IQR estimator of [Chernozhukov and Hansen \(2006\)](#) based on a grid search over 500 points. IQR provides a very robust benchmark. However, due to the use of grid search, it is computationally expensive. In our empirical Monte Carlo study in Section 8, we find that, with 10,000 observations and one endogenous regressor, IQR is nine times slower than the contraction algorithm and 23 times slower than the root-finding algorithm. With two endogenous regressors, the computational advantages of our procedures are even more pronounced. IQR based on a grid search over 100×100 points is 149 times slower than the contraction algorithm and 84 times slower than the nested root-finding algorithm.

Figure 2 displays the estimates of $\theta_D(\tau)$ for $\tau \in \{0.15, 0.20, \dots, 0.85\}$. We can see that all estimation algorithms yield very similar results. We also note that the contraction algorithm converges for all quantile levels considered.

Figure 3 depicts pointwise 95% confidence intervals for the proposed estimators obtained using the empirical bootstrap described in Section 6.2 with 500 replications. We can see that the resulting confidence intervals are very similar for both algorithms and do not include zero at any quantile level considered.

8. EMPIRICAL MONTE CARLO STUDY

In this section, we assess the practical performance of our estimation algorithms in an empirical Monte Carlo study based on the application in Section 7.

8.1 An application-based DGP

We consider DGPs which are based on the empirical application of Section 7.¹⁹ We focus on a simplified setting with only two exogenous covariates: income and age. The covari-

¹⁹The construction of our DGPs is inspired by the application-based DGPs in [Kaplan and Sun \(2017\)](#).

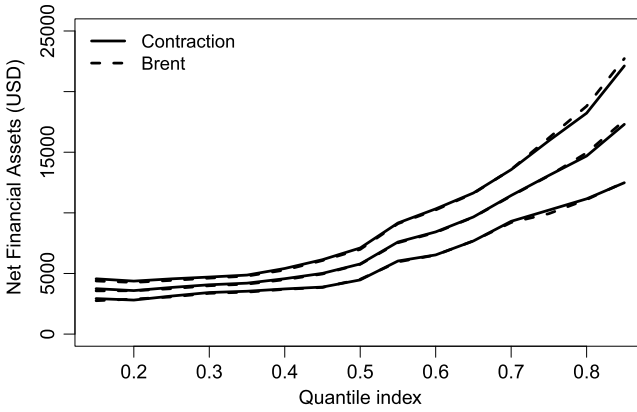


FIGURE 3. Pointwise 95% bootstrap confidence intervals.

ates are drawn from their joint empirical distribution. The instrument Z_i is generated as Bernoulli(\bar{Z}), where \bar{Z} is the mean of the instrument in the data. We then generate the endogenous variable as $D_i = Z_i \cdot 1\{0.6 \cdot V_i < U_i\}$, where $U_i \sim \text{Uniform}(0, 1)$ and $V_i \sim \text{Uniform}(0, 1)$ are independent disturbances. The DGP for D_i roughly matches the joint empirical distribution of (D_i, Z_i) . The outcome variable Y_i is generated as

$$Y_i = X_i' \theta_X(U_i) + D_i \theta_D(U_i) + G^{-1}(U_i).$$

The coefficient $\theta_X(\cdot)$ is constant and equal to the IQR median estimate in the empirical application. $\theta_D(U_i) = 5000 + U_i \cdot 10,000$ is chosen to match the increasing shape of the estimated conditional quantile treatment effects in Figure 2. $G^{-1}(\cdot)$ is the quantile function of a recentered Gamma distribution fitted to the distribution of the IQR residuals at the median. To investigate the performance of our procedure with more than one endogenous variable, we add a second endogenous regressor:

$$Y_i = X_i' \theta_X(U_i) + D_i \theta_D(U_i) + D_{2,i} \theta_{D,2}(U_i) + G^{-1}(U_i),$$

where we set $\theta_{D,2}(U_i) = 10,000$. The second endogenous variable is generated as

$$D_{2,i} = 0.8 \cdot Z_{2,i} + 0.2 \cdot \Phi^{-1}(U_i)$$

and the second instrument is generated as $Z_{2,i} \sim N(0, 1)$. We set $N = 9913$ as in the empirical application.

8.2 Estimation algorithms

We assess and compare several different algorithms, all of which are based on the dynamical system \hat{M} .²⁰ For models with one endogenous variable, we consider a contraction algorithm and a root-finding algorithm based on Brent’s method. For models with

²⁰We do not explore algorithms based on \hat{K} because, as discussed in Section 5, the algorithms based on \hat{M} have advantages in terms of the choice of starting values (for the contraction algorithm) and dimension reduction (for the root-finding algorithm).

TABLE 1. Algorithms.

<i>One endogenous variable</i>	
Algorithm	R-Function
Contraction algorithm	
Root-finding algorithm	uniroot (R Core Team (2020))
Profiling algorithm	uniroot (R Core Team (2020))
IQR	
<i>Two endogenous variables</i>	
Algorithm	R-Function
Contraction algorithm	
Nested root-finding algorithm	uniroot (R Core Team (2020))
Root-finding algorithm (implemented as optimizer)	optim_sa (Husmann, Lange, and Spiegel (2017))
Nested profiling algorithm	uniroot (R Core Team (2020))
IQR (implementation: p. 132 in Chernozhukov, Hansen, and Wüthrich (2017))	

two endogenous variables, we consider a contraction algorithm, a nested root-finding algorithm based on Brent's method, and a root-finding algorithm implemented as a minimization problem based on simulated annealing (SA).²¹ For all estimators, we use 2SLS estimates as starting values. We compare the results of our algorithms to (nested) profiling estimators based on Brent's method and to IQR with a grid search over 500 points (one endogenous regressor) and 100×100 points (two endogenous regressors), which serves as a slow but very robust benchmark. Table 1 presents an overview of the algorithms.

8.3 Results

Here, we describe the computational performance of the different procedures and the finite sample performance of our bootstrap inference procedure. In Appendix D.1, we further investigate the finite sample bias and root mean squared error (RMSE) of the different methods. We find that the proposed estimation algorithms perform well and exhibit a similar bias and RMSE. Moreover, the finite sample properties of the estimators based on the contraction and root-finding algorithms are comparable to the profiling estimators and IQR. This shows that the computational advantages of our algorithms do not come at a cost in terms of the finite sample performance. Appendix D.2 presents additional simulation evidence, demonstrating that our algorithms perform well and remain computationally tractable with more than two endogenous regressors.²²

²¹We have also explored algorithms based on Newton–Raphson-type root-finders. These algorithms are, in theory, up to an order of magnitude faster than the contraction algorithm and the nested algorithms, but, unlike the other algorithms considered here, require an approximation to the Jacobian and are not very robust to the choice of starting values. We therefore do not report the results here.

²²Specifically, we present simulation results based on the DGP used in this section, augmented with an additional endogenous regressor, generated as $D_{3,i} = 0.8 \cdot Z_{3,i} + 0.2 \cdot \Phi^{-1}(U_i)$, where $Z_{3,i} \sim N(0, 1)$.

TABLE 2. Computation time, 401(k) DGP with one endogenous regressor.

N	Contr	Brent	Profil	InvQR
1000	0.10	0.03	0.03	0.34
5000	0.44	0.13	0.12	2.44
10,000	0.75	0.28	0.30	6.48

Note: The table reports average computation time in seconds at $\tau = 0.5$ over 100 simulation repetitions based on the DGP described in the main text. Contr: contraction algorithm; Brent: root-finding algorithm based on Brent's method; Profil: profiling estimator based on Brent's method; InvQR: inverse QR with grid search over 500 grid points. We use 2SLS estimates as starting values.

Tables 2–3 show the average computation time (in seconds) for estimating the model with one and two endogenous variables for different sample sizes. All computations were carried out on a standard desktop computer with a 3.2 GHz Intel Core i5 processor and 8 GB RAM.

With one endogenous regressor, the contraction algorithm and the root-finding algorithm based on Brent's method are computationally more efficient than IQR. Specifically, the root-finding algorithm based on Brent's method is 11 to 23 times faster than IQR and the contraction algorithm is 4 to 9 times faster. The root-finding algorithm is as fast as the profiling estimator and about twice as fast as the contraction algorithm.²³

The computational advantages of our algorithms become more pronounced with two endogenous variables. Table 3 shows that IQR's average computation times are around two orders of magnitude slower than those of our preferred procedures. Specifically, the nested root-finding algorithm is 84 to 134 times faster than IQR, while the contraction algorithm is 149 to 308 times faster. This is as expected since, due to the use of grids, IQRs become computationally impractical in our setting whenever the number of

TABLE 3. Computation time, 401(k) DGP with two endogenous regressors.

N	Contr	NestBr	SimAnn	Profil	InvQR
1000	0.18	0.41	3.13	0.77	55.40
5000	1.71	2.67	19.05	4.39	282.64
10,000	4.50	7.99	54.95	12.50	672.00

Note: The table reports average computation time in seconds at $\tau = 0.5$ over 100 simulation repetitions based on the DGP described in the main text. Contr: contraction algorithm; NestBr: nested algorithm based on Brent's method; SimAnn: simulated annealing based optimization algorithm; Profil: nested profiling estimator based on Brent's method; InvQR: inverse QR with grid search over 100×100 grid points. We use 2SLS estimates as starting values.

²³Note that the computational speed of the contraction algorithm depends on $|\hat{J}_M|$, and thus, will differ across applications.

TABLE 4. Coverage, 401(k) DGP with one endogenous regressor.

τ	$1 - \alpha = 0.95$		$1 - \alpha = 0.9$	
	Contr	Brent	Contr	Brent
0.15	0.96	0.96	0.93	0.91
0.25	0.95	0.95	0.91	0.92
0.50	0.95	0.95	0.90	0.91
0.75	0.95	0.95	0.90	0.89
0.85	0.94	0.94	0.89	0.89

Note: Monte Carlo simulation with 1000 repetitions as described in the main text. Contr: contraction algorithm; Brent: root-finding algorithm based on Brent's method. We use 2SLS estimates as starting values.

endogenous variables exceeds two or three.²⁴ The contraction algorithm is almost twice as fast as the nested algorithm, which, in turn, is almost twice as fast as the profiling estimator. Finally, the minimization algorithm based on SA is about one order of magnitude slower than the contraction algorithm, while still being an order of magnitude faster than IQR.

Finally, we analyze the finite sample properties of our bootstrap inference procedure for making inference on θ_D in the model with a single endogenous variable. Table 4 shows the empirical coverage probabilities of bootstrap confidence intervals based on the contraction algorithm and the root-finding algorithm based on Brent's method. Both methods demonstrate an excellent performance and exhibit coverage rates that are very close to the respective nominal levels.

9. CONCLUSION

The main contribution of this paper is to develop computationally convenient and easy-to-implement estimation algorithms for IVQR models. Our key insight is that the nonsmooth and nonconvex IVQR estimation problem can be decomposed into a sequence of much more tractable convex QR problems, which can be solved very quickly using well-established methods. The proposed algorithms are particularly well suited if the number of exogenous variables is large and the number of endogenous variables is moderate as in many empirical applications.

An interesting avenue for further research is to investigate weak identification robust inference within the decentralized model. One may, for example, write the (rescaled) sample fixed-point restriction as $\sqrt{N}(I_{d_X+d_D} - \hat{K})(\theta) = s_N(\theta) + \mathbb{W}(\theta) + r_N(\theta)$, where

²⁴Our implementation of IQR with two endogenous variables is inherently slower than the implementation with one endogenous variable, even when the number of grid points is the same. First, there is an additional covariate in the underlying QRs (the second instrument). Second, with one endogenous variable, we choose the grid value that minimizes the absolute value of the coefficient on the instrument. By contrast, with two endogenous regressors, we choose the grid point which minimizes a quadratic form based on the inverse of the estimated QR variance covariance matrix as suggested in Chernozhukov, Hansen, and Wüthrich (2017), which requires an additional computational step.

$s_N(\theta) = \sqrt{N}(I_{d_X+d_D} - K)(\theta)$, \mathbb{W} is a Gaussian process, and r_N is an error that tends to zero uniformly. This paper assumes that $s_N(\theta^*) = 0$ uniquely, and outside $N^{-1/2}$ -neighborhoods of θ^* , $s_N(\theta)$ diverges and dominates \mathbb{W} . For a one-dimensional fixed-point problem, this requires the BR map to be bounded away from the 45-degree line outside any $N^{-1/2}$ -neighborhood of the fixed point. However if s_N fails to dominate \mathbb{W} over a substantial part of the parameter space, one would end up with weak identification.²⁵ How to conduct robust inference in such settings is an interesting question, which we leave for future research.

Finally, we note that, while we study the performance of the proposed algorithms separately, our reformulation and the resulting algorithms are potentially very useful when combined with other existing procedures. For instance, one could choose starting values using an initial grid search over a coarse grid and then apply a fast contraction algorithm.

REFERENCES

- Abadie, A., J. Angrist, and G. Imbens (2002), “Instrumental variable estimates of the effect of subsidized training on the quantile of trainee earnings.” *Econometrica*, 70 (1), 91–117. [446]
- Ambrosetti, A. and G. Prodi (1995), *A Primer of Nonlinear Analysis*. Cambridge Studies in Advanced Mathematics, Vol. 34. Cambridge University Press, Cambridge. [456]
- Andrews, I. and A. Mikusheva (2016), “Conditional inference with a functional nuisance parameter.” *Econometrica*, 84 (4), 1571–1612. [445, 471]
- Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012), “Estimating spillovers using panel data, with an application to the classroom.” *Quantitative Economics*, 3 (3), 421–470. [446]
- Belloni, A., V. Chernozhukov, I. Fernandez-Val, and C. Hansen (2017), “Program evaluation and causal inference with high-dimensional data.” *Econometrica*, 85 (1), 233–298. [446]
- Benjamin, D. (2003), “Does 401(k) eligibility increase saving?: Evidence from propensity score subclassification.” *Journal of Public Economics*, 87, 1259–1290. [464]
- Brent, R. P. (1971), “An algorithm with guaranteed convergence for finding a zero of a function.” *The Computer Journal*, 14 (4), 422–425. [460]
- Chen, L.-Y. and S. Lee (2018), “Exact computation of GMM estimators for instrumental variable quantile regression models.” *Journal of Applied Econometrics*, 33 (4), 553–567. [445]
- Chen, M., I. Fernandez-Val, and M. Weidner (2014), “Nonlinear panel models with interactive effects.” [arXiv:1412.5647v1](https://arxiv.org/abs/1412.5647v1). [447]

²⁵Andrews and Mikusheva (2016) studied weak identification robust inference methods in models characterized by moment restrictions.

- Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014), “Local identification of non-parametric and semiparametric models.” *Econometrica*, 82 (2), 785–809. [455]
- Chen, X. and D. Pouzo (2009), “Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals.” *Journal of Econometrics*, 152 (1), 46–60. [446]
- Chen, X. and D. Pouzo (2012), “Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals.” *Econometrica*, 80 (1), 277–321. [446]
- Chernozhukov, V. and C. Hansen (2004), “The effects of 401(k) participation on the wealth distribution: An instrumental quantile regression analysis.” *The Review of Economics and Statistics*, 86 (3), 735–751. [445, 464]
- Chernozhukov, V. and C. Hansen (2005), “An IV model of quantile treatment effects.” *Econometrica*, 73 (1), 245–261. [443, 447, 448]
- Chernozhukov, V. and C. Hansen (2006), “Instrumental quantile regression inference for structural and treatment effects models.” *Journal of Econometrics*, 132, 491–525. [443, 444, 445, 448, 456, 457, 463, 466]
- Chernozhukov, V. and C. Hansen (2008), “Instrumental variable quantile regression: A robust inference approach.” *Journal of Econometrics*, 142 (1), 379–398. [445]
- Chernozhukov, V. and C. Hansen (2013), “Quantile models with endogeneity.” *Annual Review of Economics*, 5 (1), 57–81. [445, 446, 448]
- Chernozhukov, V., C. Hansen, and M. Jansson (2009), “Finite sample inference for quantile regression models.” *Journal of Econometrics*, 152 (2), 93–103. [445]
- Chernozhukov, V., C. Hansen, and K. Wüthrich (2017), “Instrumental variable quantile regression.” In *Handbook of Quantile Regression* (V. Chernozhukov, X. He, R. Koenker, and L. Peng, eds.), 119–143, CRC Press, Boca Raton. [446, 460, 468, 470]
- Chernozhukov, V. and H. Hong (2003), “An MCMC approach to classical estimation.” *Journal of Econometrics*, 115 (2), 293–346. [445]
- Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007), “Instrumental variable estimation of nonseparable models.” *Journal of Econometrics*, 139 (1), 4–14. [446]
- Chesher, A. (2003), “Identification in nonseparable models.” *Econometrica*, 71 (5), 1405–1441. [446]
- de Castro, L., A. F. Galvao, D. M. Kaplan, and X. Liu (2019), “Smoothed GMM for quantile models.” *Journal of Econometrics*, 213 (1), 121–144. *Annals: In Honor of Roger Koenker*. [445]
- D’Haultfoeulle, X. and P. Février (2015), “Identification of nonseparable triangular models with discrete instruments.” *Econometrica*, 83 (3), 1199–1210. [446]
- Dominitz, J. and R. P. Sherman (2005), “Some convergence theory for iterative estimation procedures with an application to semiparametric estimation.” *Econometric Theory*, 21 (04), 838–863. [461]

- Engen, E. M., W. G. Gale, and J. K. Scholz (1996), “The illusory effects of saving incentives on saving.” *Journal of Economic Perspectives*, 10 (4), 113–138. [464]
- Frandsen, B. R., M. Frölich, and B. Melly (2012), “Quantile treatment effects in the regression discontinuity design.” *Journal of Econometrics*, 168 (2), 382–395. [446]
- Frölich, M. and B. Melly (2013), “Unconditional quantile treatment effects under endogeneity.” *Journal of Business & Economic Statistics*, 31 (3), 346–357. [446]
- Fudenberg, D. and D. K. Levine (2007), *The Theory of Learning in Games*. MIT Press, Cambridge. [453]
- Gagliardini, P. and O. Scaillet (2012), “Nonparametric instrumental variable estimation of structural quantile effects.” *Econometrica*, 80 (4), 1533–1562. [446]
- Gale, D. and H. Nikaido (1965), “The Jacobian matrix and global univalence of mappings.” *Mathematische Annalen*, 159 (2), 81–93. [456]
- Galor, O. (2007), *Discrete Dynamical Systems*. Springer Science & Business Media, Berlin, Heidelberg, New York, NY. [455]
- Guimaraes, P. and P. Portugal (2010), “A simple feasible procedure to fit models with high-dimensional fixed effects.” *Stata Journal*, 10 (4), 628–649. [446]
- Horowitz, J. L. and S. Lee (2007), “Nonparametric instrumental variables estimation of a quantile regression model.” *Econometrica*, 75 (4), 1191–1208. [446]
- Husmann, K., A. Lange, and E. Spiegel (2017), *The R Package Optimization: Flexible Global Optimization With Simulated-Annealing*. [468]
- Imbens, G. W. and J. D. Angrist (1994), “Identification and estimation of local average treatment effects.” *Econometrica*, 62 (2), 467–475. [446]
- Imbens, G. W. and W. K. Newey (2009), “Identification and estimation of triangular simultaneous equations models without additivity.” *Econometrica*, 77 (5), 1481–1512. [446]
- Jun, S. J. (2008), “Weak identification robust tests in an instrumental quantile model.” *Journal of Econometrics*, 144 (1), 118–138. [445]
- Jun, S. J. (2009), “Local structural quantile effects in a model with a nonseparable control variable.” *Journal of Econometrics*, 151 (1), 82–97. [446]
- Kaido, H. and K. Wüthrich (2021), “Supplement to ‘Decentralization estimators for instrumental variable quantile regression models’.” *Quantitative Economics Supplemental Material*, 12, <https://doi.org/10.3982/QE1440>. [447]
- Kaplan, D. M. and Y. Sun (2017), “Smoothed estimation equations for instrumental variables quantile regression.” *Econometric Theory*, 33 (1), 105–157. [445, 466]
- Koenker, R. (2017), “Computational methods for quantile regression.” In *Handbook of Quantile Regression* (V. Chernozhukov, X. He, R. Koenker, and L. Peng, eds.), 55–67, CRC Press, Boca Raton. [450, 458]

- Koenker, R. (2020), “quantreg: Quantile Regression R package version 5.55.” [458]
- Koenker, R. and G. J. Bassett (1978), “Regression quantiles.” *Econometrica*, 46 (1), 33–50. [443]
- Koenker, R. W. and V. D’Orey (1987), “Algorithm AS 229: Computing regression quantiles.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36 (3), 383–393. [458]
- Koenker, R. W. and V. D’Orey (1994), “Remark AS R92: A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43 (2), 410–414. [458]
- Lee, J. and K. Seo (2015), “A computationally fast estimator for random coefficients logit demand models using aggregate data.” *The RAND Journal of Economics*, 46 (1), 86–102. [447]
- Lee, S. (2007), “Endogeneity in quantile regression models: A control function approach.” *Journal of Econometrics*, 141 (2), 1131–1158. [446]
- Li, S. and T. Basar (1987), “Distributed algorithms for the computation of noncooperative equilibria.” *Automatica*, 23 (4), 523–533. [453]
- Ma, L. and R. Koenker (2006), “Quantile regression methods for recursive structural equation models.” *Journal of Econometrics*, 134 (2), 471–506. [446]
- Machado, J. A. and J. Santos Silva (2019), “Quantiles via moments.” *Journal of Econometrics*, 213 (1), 145–173. Annals: In Honor of Roger Koenker. [446]
- Marra, G. and R. Radice (2013), “Estimation of a regression spline sample selection model.” *Computational Statistics & Data Analysis*, 61, 158–173. [446]
- Melly, B. and K. Wüthrich (2017), “Local quantile treatment effects.” In *Handbook of Quantile Regression* (V. Chernozhukov, X. He, R. Koenker, and L. Peng, eds.), 145–164, CRC Press, Boca Raton. [446]
- Moon, H. R. and M. Weidner (2015), “Linear regression for panel with unknown number of factors as interactive fixed effects.” *Econometrica*, 83 (4), 1543–1579. [447]
- Portnoy, S. and R. Koenker (1997), “The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators.” *Statistical Science*, 12 (4), 279–300. [458]
- Poterba, J. M., S. F. Venti, and D. A. Wise (1994), “401(k) plans and tax-deferred saving.” In *Studies in the Economics of Aging* (D. A. Wise, ed.), 105–142, University of Chicago Press, Chicago and London. [464]
- Poterba, J. M., S. F. Venti, and D. A. Wise (1995), “Do 401(k) contributions crowd out other personal saving?” *Journal of Public Economics*, 58 (1), 1–32. [464]
- Poterba, J. M., S. F. Venti, and D. A. Wise (1998), “Personal retirement saving programs and asset accumulation: Reconciling the evidence.” In *Frontiers in the Economics of Aging* (D. A. Wise, ed.), 23–124, University of Chicago Press, Chicago and London. [464]

- Pouliot, G. A. (2019), “Instrumental variables quantile regression with multivariate endogenous variable.” Working Paper, University of Chicago. [445, 446]
- R Core Team (2020), *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [468]
- Rothenberg, T. J. (1971), “Identification in parametric models.” *Econometrica*, 39 (3), 577–591. [455]
- Rubin, D. B. (1974), “Estimating causal effects of treatment in randomized and nonrandomized studies.” *Journal of Educational Psychology*, 66 (5), 688–701. [447]
- Sakov, A. and P. J. Bickel (2000), “An Edgeworth expansion for the m out of n bootstrapped median.” *Statistics & Probability Letters*, 49 (3), 217–223. [463]
- Smyth, G. K. (1996), “Partitioned algorithms for maximum likelihood and other non-linear estimation.” *Statistics and Computing*, 6 (3), 201–216. [446, 453]
- Torgovitsky, A. (2015), “Identification of nonseparable models using instruments with small support.” *Econometrica*, 83 (3), 1185–1197. [446]
- Weisberg, S. and A. H. Welsh (1994), “Adapting for the missing link.” *The Annals of Statistics*, 22 (4), 1674–1700. [446]
- Wüthrich, K. (2019), “A closed-form estimator for quantile treatment effects with endogeneity.” *Journal of Econometrics*, 210 (2), 219–235. [446]
- Wüthrich, K. (2020), “A comparison of two quantile models with endogeneity.” *Journal of Business & Economic Statistics*, 38 (2), 443–456. [446]
- Zhu, Y. (2018), “k-step correction for mixed integer linear programming: A new approach for instrumental variable quantile regressions and related problems.” arXiv:1805.06855v2. [445]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 11 September, 2019; final version accepted 13 November, 2020; available online 21 December, 2020.