Identification and estimation of nonparametric panel data regressions with measurement error

Daniel Wilhelm

Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error[*]

Daniel Wilhelm[†]

UCL and CeMMAP

June 28 , 2015

Abstract

This paper provides a constructive argument for identification of nonparametric panel data models with measurement error in a continuous explanatory variable. The approach point identifies all structural elements of the model using only observations of the outcome and the mismeasured explanatory variable; no further external variables such as instruments are required. In the case of two time periods, restricting either the structural or the measurement error to be independent over time allows past explanatory variables or outcomes to serve as instruments. Time periods have to be linked through serial dependence in the latent explanatory variable, but the transition process is left nonparametric. The paper discusses the general identification result in the context of a nonlinear panel data regression model with additively separable fixed effects. It provides a nonparametric plug-in estimator, derives its uniform rate of convergence, and presents simulation evidence for good performance in finite samples.

1 Introduction

A vast amount of empirical work in economics, especially in microeconomics, recognizes the possible presence of measurement error (ME) in relevant data sets and that failure to account for it may result in misleading estimates and inference.[1] These concerns deserve particular attention in panel data models. Economists view panel data as very appealing because they offer opportunities to flexibly deal with unobserved, individual-specific heterogeneity in economic behavior. In one form or another, panel data estimators combine variables from different time periods to remove ("difference out") those individual-specific components, typically resulting in transformed variables with amplified ME relative to the original variables. In this sense, ME poses an even more severe challenge to panel data models than to cross-sectional ones.

This paper shows the possibly nonlinear relationship between an individual's outcome and unobserved true explanatory variable can be recovered from only observed variables within the panel model (i.e. outcomes and mismeasured covariates). Specifically, consider the nonlinear regression model with additively separable fixed effects,

$$Y_{it} = m(X_{it}^*) + \alpha_i + \varepsilon_{it}, \tag{1}$$

which in the subsequent discussion, serves as the leading example to which the new approach can be applied. The dependent variable Y_{it} denotes an individual i's outcome in time period t, α_i an individual-specific fixed effect, which may arbitrarily depend on a continuous regressor X_{it}^*, and ε_{it} the regression error, also referred to as the structural error. Instead of the true explanatory variable X_{it}^*, only the mismeasured variable X_{it} is observed. The ME $\eta_{it} := X_{it} - X_{it}^*$ is assumed classical in the sense that it does not depend on the true regressor, but nonclassical features like temporal dependence in the ME are allowed. The structural relationship between Y_{it} and X_{it}^*, in (1) represented by the unknown function $m(x^*)$, is the object of interest. The identification approach in this paper provides explicit formulae (in terms of observable variables) for $m(x^*)$, for the distribution of the ME, and for the distribution of the true regressor jointly over all time periods. The identification result does *not* require any observed variables other than outcomes and observed regressors for two or four time periods, depending on whether additive fixed effects are present. The idea consists of taking past observed regressors or past outcomes as instrumental variables. Consequently, the main identifying assumptions resemble those of standard linear instrumental variable models: (i) outcomes are determined by the true, not the mismeasured, covariates; (ii) different time periods are linked

---

[1]See the excellent handbook chapter by Bound, Brown, and Mathiowetz (2001).

through serial dependence in the true covariates (relevance condition); and (iii) past mismeasured covariates or past outcomes do not determine present outcomes, conditional on present, true covariates (exclusion restriction). Depending on whether covariates or outcomes are used as instruments, I show that the exclusion restriction requires either the structural error or the ME to be independent over time, respectively. The constructive nature of the approach suggests nonparametric estimation of all model components by simply substituting nonparametric estimators into the population formulae and choosing values for smoothing parameters. The resulting estimator can be computed based only on basic operations such as matrix multiplication, matrix inversion, and discrete Fourier transforms, which are carried out efficiently on modern computers, and does not require any optimization routines.

As an example in which the methods developed in this paper may lead to interesting new conclusions, consider the relationship between investment and Tobin's q in Lucas and Prescott (1971)-type investment models. The firm's profit maximization problem leads to first-order conditions of the form

$$\frac{I_{it}}{K_{it}} = m(q_{it}^*) + \alpha_i + \varepsilon_{it},$$

where I_{it} denotes investment and K_{it} capital stock of firm i at time t. The unobserved regressor q_{it}^*, also called the shadow value of capital or marginal q, is the Lagrange multiplier for the evolution equation of the capital stock. The ratio of book to market value of the firm's capital stock, q_{it}, also called average q, is a popular proxy for q_{it}^*. To estimate the model, empirical work typically imposes one or both of the following two assumptions: (i) firms face a quadratic adjustment cost function and (ii) unobserved marginal q is equal to observed average q. Assumption (i) leads to a linear function m, but has no economic foundation and has been argued to be likely too strong (Barnett and Sakellaris (1998)). Assumption (ii) eliminates the measurement error problem but imposes strong conditions on the economic framework (Hayashi (1982)) and have also been argued to be unrealistic (Erickson and Whited (2000), Almeida, Campello, and Galvao (2010), Erickson and Whited (2012)). Since ME and nonlinearities can manifest themselves in similar ways (Chesher (1991)), the ability to analyze the investment model without imposing either of (i) and (ii) is important. In the investment literature, accounting for ME while allowing for the presence of nonlinearities in m is a challenge because of the lack of external instruments. However, large firm-level panel data sets are readily available and therefore render the approach taken in this paper immediately applicable. The author is pursuing a thorough study in this direction in a separate paper.

Other examples are the estimation of Engel functions from household panel data (Aas-

ness, Biørn, and Skjerpen (1993)), the estimation of production functions (Olley and Pakes (1996)), studies of returns to research and development performed in private firms (Hall, Jaffe, and Trajtenberg (2005)), and analyzing the technology of skill formation (Cunha, Heckman, and Schennach (2010)). Since the first draft, the identification argument of this paper has also been used and extended for the study of nonlinearities in the dynamics of household income (Arellano (2014); Arellano, Blundell, and Bonhomme (2014)).

Existing work on the treatment of ME in panel data models is scarce and exclusively[2] focuses on linear specifications. In the well-known approach by Griliches and Hausman (1986), the linearity facilitates the derivation of explicit formulae for the biases of the first difference and the within estimator. In both bias expressions, the variance of the ME is the only unknown, so, from the two different estimators, one can substitute out the unknown variance and calculate a ME-robust estimator. Clearly, such an approach cannot be expected to carry through to nonlinear models. More recent approaches such as Holtz-Eakin, Newey, and Rosen (1988), Biørn (2000), Buonaccorsi, Demidenko, and Tosteson (2000), Wansbeek (2001), Xiao, Shao, and Palta (2008), Galvao Jr. and Montes-Rojas (2009), Shao, Xiao, and Xu (2011), and Komunjer and Ng (2011) similarly rely heavily on a linear specification and cannot identify or estimate nonlinear models such as (1).

The ME literature for nonlinear cross-sectional models is extensive. Most identification arguments assume the availability of instruments, repeated measurements, or auxiliary data. Chen, Hong, and Nekipelov (2011) and Schennach (2013) review this stream of the literature and provide numerous references. The existing instrumental variable approaches require the instrument to predict the true covariate in a certain sense, whereas the true value predicts the mismeasured covariate.[3] This asymmetry between assumptions on the mismeasured and the instrumental variable conflicts with the structure of conventional panel data models such as (1) when mismeasured covariates or outcomes are the only candidates for instruments. Looking for suitable instruments excluded from the panel model is not a solution either because the motivation for using panel data oftentimes lies in the desire to deal with endogeneity when external variation, for instance in the form of instruments, is unavailable. Repeated measurements or auxiliary data for panel data sets are difficult, if not impossible, to acquire[4], so that approaches based on their availability appear to be of little use for panel models as well. In conclusion, the existing

---

[2]Part of the statistic literature has approached the problem by assuming the ME is classical and its distribution known (see Carroll, Ruppert, Stefanski, and Crainiceanu (2006) and references therein). In this special case, which is unrealistic in most economic applications, identification is straightforward.

[3]Hu and Schennach (2008) is an exception; see the discussion below.

[4]Card (1996) describes an interesting and fortunate exception, but he considers misclassification of a binary regressor, whereas this paper is concerned with continuous covariates.

4

approaches for nonlinear cross-sectional models with ME either impose assumptions on instruments incompatible with panel data models such as (1) or require auxiliary data that is typically not available. In contrast, the present paper provides conditions for identification based only on observations from within the panel. A separate stream of the literature combines independence assumptions with information in higher-order moments of observed variables; see Schennach and Hu (2013) and the many references therein.

Hu and Schennach (2008) make an interesting theoretical contribution by providing a powerful operator diagonalization result that can be used to identify a large class of models with latent variables. The cost of their generality are high-level assumptions and an identification argument that is not constructive. Furthermore, in regression models, their assumptions rule out homoskedastic errors with a regression function that is equal at two points. On the contrary, the new approach of this paper is constructive, leads to assumptions that are easily interpretable in a panel data context, and allows the treatment of time dependence in the ME. In general, neither their approach nor mine nests the other.

The present paper also relates to the statistical literature on ill-posed inverse problems that has gained recent interest in the econometrics community, for example, Newey and Powell (2003), Hall and Horowitz (2005), Horowitz and Lee (2007), Blundell, Chen, and Kristensen (2007), Carrasco, Florens, and Renault (2007), Horowitz (2009), Darolles, Fan, Florens, and Renault (2011), Horowitz (2011), and Chetverikov and Wilhelm (2015). Part of the identification argument of this paper consists of solving two ill-posed inverse problems similar to those arising in nonparametric instrumental variable models considered in these papers. My approach implements Blundell, Chen, and Kristensen (2007)'s series estimator as an input to a second-stage estimator of the actual quantity of interest.

Porter (1996), Henderson, Carroll, and Li (2008), Evdokimov (2010), Qian and Wang (2012), Lee (2014), and Fève and Florens (2014) discuss nonparametric identification and estimation of panel data models without ME, but their models are otherwise similar to the leading example in this paper, namely, nonparametric panel data regressions with additively separable fixed effects such as (1).

2 Identification

2.1 A General Instrumental Variable Identification Result

Consider a panel data model with an individual's outcome Y_t, which can be discrete or continuous, and a continuous explanatory variable X_t^*. Both variables are indexed by the time period t, but for the remainder of the paper, I omit the individual's subscript i. The

relationship between these two variables, subsequently referred to as the structural relationship, is the object of interest in this paper. Assume the researcher does not observe X_t^* directly, but only a mismeasured version of it, denoted by X_t. The difference between the two, $\eta_t := X_t - X_t^*$, is referred to as ME. Standard approaches such as nonparametric regression or nonparametric instrumental variables regression cannot recover features of the structural relationship because the independent variable is unobserved. In this section, I provide assumptions restricting the joint distribution of (Y_t, X_t^*, X_t) so that, in the population, the structural relationship can be uniquely determined from observable quantities. Although the paper's motivation is to treat ME in panel data models, the approach to identification uses past observed variables in a general way so other instruments can replace these variables. Therefore, the identification result may be of more general interest, even in cross-sectional models; see Remark 2 below.

For now suppose only two time periods are available and no individual-specific effects or any other perfectly measured covariates are present. I discuss such extensions in the next subsection. Denote by Y (X, X^* and η) the vectors (matrices) stacking the corresponding variables for individual time periods.[5]

Assumption ID 1. *The distributions of X, X^*, and η admit densities f_X, f_{X^*}, and f_η with respect to Lebesgue measure on $\mathbb{R}^p$ and have support equal to $\mathbb{R}^p$.*

The first assumption restricts the latent and observed independent variables to be continuously distributed with infinite support.[6] The outcome, however, is allowed to be discrete or continuous.

Assumption ID 2. **(i)** $\eta \perp X^*$ *and* $E\eta = 0$ *and* **(ii)** *the characteristic function of η is nonzero everywhere.*

The first part of Assumption ID 2 specifies that the ME is independent of the true value and has mean zero. This assumption is strong, but standard in the ME literature dealing with continuous mismeasured variables. While it would be possible to allow for some dependence between the ME and the true value, imposing independence leads to identifying assumptions that are easy to interpret in the context of a panel structure, a constructive identification argument, and, thus, simple nonparametric plugin estimators for all components of the model. Notice that the ME is "non-classical" in the sense

---

[5]By a slight abuse of notation, when the dimension of the covariates is larger than one, I may refer to the distribution of the random matrices X, X^*, or η, which should be interpreted as the distribution of the random vectors $vec(X)$, $vec(X^*)$, or $vec(\eta)$, respectively.

[6]Relaxing the assumption of unbounded support along the lines of Carrasco and Florens (2011) appears possible.

that part (i) of the assumption does not restrict the temporal dependence of the ME. The second part of Assumption ID 2 is common in the deconvolution literature; see Carroll, Ruppert, Stefanski, and Crainiceanu (2006), Fan (1991), Fan and Truong (1993), Li and Vuong (1998) and Schennach (2004a, 2007). It excludes, for example, uniform and triangular distributions, but most other standard distributions such as the Normal, t, χ^2, gamma, and double exponential satisfy the requirement.

Assumption ID 3. (i) *There exist $B < \infty$ and $\beta \in \mathbb{R}^p_+$ such that*

$$\sup_{y_2 \in \mathbb{Y}_2} \left| \frac{\partial^p}{\partial x_2^{*(1)} \cdots \partial x_2^{*(p)}} F_{Y_2|X_2^*}(y_2|x_2^*) \right| \leq B \prod_{k=1}^p \left(1 + \left| x_2^{*(k)} \right| \right)^{-3-\beta_k}$$

for all $x_2^* = (x_2^{*(1)}, \ldots, x_2^{*(p)})' \in \mathbb{R}^p$, where $|\cdot|$ denotes the Euclidean norm; **(ii)** *f_η is bounded.*

In economic models, Assumption ID 3 is a relatively weak assumption on the conditional cdf of Y_2 given X_2^* as the conditioning argument becomes large. It rules out rather pathological cases in which the cdf oscillates too much as $|x_2^*| \to \infty$. The condition guarantees that after appropriate centering, the Fourier transform of the conditional cdf is an ordinary function even though the conditional cdf is not absolutely integrable in its conditioning argument; see Schennach (2008).

Definition 1. *Let the notation $x^{(k)}$ refer to the k-th element of a vector x. For some space $\bar{\mathcal{G}} \subseteq L^2(f_{X_2^*})$ that contains $F_{Y_2|X_2^*}(y_2|\cdot)$ for every $y_2 \in \mathbb{Y}_2$, define the set of functions $\mathcal{G} := \mathcal{G}_1 \cup \mathcal{G}_2$ with*

$$\mathcal{G}_1 := \left\{ h \in \bar{\mathcal{G}} : \ 0 \leq h(x) \leq 1 \quad \forall x \in \mathbb{R}^p \right\},$$
$$\mathcal{G}_2 := \left\{ h \in \bar{\mathcal{G}} : \ \exists k \in \{1, \ldots, p\}, \ \bar{h} \in \mathcal{G}_1 : \ h(x) = x^{(k)}\bar{h}(x) \quad \forall x \in \mathbb{R}^p \right\}.$$

The set $\mathcal{G}$ contains two types of functions: the first component, $\mathcal{G}_1$, is the set of functions bounded between zero and one, and the second, $\mathcal{G}_2$, contains the bounded functions from $\mathcal{G}_1$ multiplied by a component of its argument.

Assumption ID 4. *The conditional distribution of X_2^* given X_1^* is $\mathcal{G}$-complete.*

$\mathcal{G}$-completeness of the distribution of X_2^* given X_1^* means that, for all functions h in $\mathcal{G}$, $E[h(X_2^*)|X_1^*] = 0$ almost surely implies $h(X_2^*) = 0$ almost surely. If X_t^* is independent over time then past or future covariates and outcomes do not contain any information about the latent variable in the present period, and the identification argument breaks down. Assumption ID 4 rules this case out. The completeness condition can be interpreted as a

nonparametric analogue of the standard rank condition in linear models. For example, if the transition process satisfies $X_2^* = \beta X_1^* + U$ with $E[UX_1^*] = 0$ then the rank condition is $E[X_1^* X_2^*]\beta = 0 \implies \beta = 0$. However, $\mathcal{G}$-completeness can accommodate much more general nonlinear specifications.

$\mathcal{G}$-completeness is a weaker restriction on the distribution of $X_2^* \mid X_1^*$ the smaller the set $\mathcal{G}$ is. In the related literature, $\mathcal{G}$ typically consists of all integrable functions or of all bounded functions, amounting to the familiar notions of completeness or bounded completeness, respectively. The type of functions $h \in \bar{\mathcal{G}}$ for which we need completeness are potential candidates $h(\cdot)$ for the desired conditional cumulative distribution function (cdf) $F_{Y_t|X_t^*}(y_t|\cdot)$. Economic theory often implies restrictions on $F_{Y_t|X_t^*}(y_t|\cdot)$ such as smoothness, monotonicity, or continuity. $\bar{\mathcal{G}}$ can then be taken as the space of functions satisfying those restrictions. One basic requirement all functions in $\bar{\mathcal{G}}$ must satisfy is Assumption ID 3(i). If $F_{Y_t|X_t^*}$ is strictly monotone in its conditioning argument (e.g., because of a strictly monotone regression function) then one may require all functions in $\bar{\mathcal{G}}$ to be asymmetric around zero. In this case, $\mathcal{G}_2$-completeness is implied by $\mathcal{G}_1$-completeness, and thus $\mathcal{G}$-completeness becomes weaker than bounded completeness. In the absence of such additional restrictions on $\bar{\mathcal{G}}$, however, $\mathcal{G}$-completeness is slightly stronger than bounded completeness and weaker than completeness for all functions bounded by polynomials (called P-completeness; D'Haultfœuille (2011)). Completeness conditions have become popular in the recent econometrics literature, and more primitive sufficient conditions are known; see Newey and Powell (2003), D'Haultfœuille (2011), Hu and Shiu (2011), and Andrews (2011). For instance, consider the transition process $X_2^* = h(X_1^*) + U$ with $U \perp X_1^*$. Under regularity assumptions, a characteristic function of U that has infinitely many derivatives and is nonzero then implies bounded completeness of $X_2^* \mid X_1^*$; see D'Haultfœuille (2011). With these restrictions on the innovations, popular time series models such as ARMA processes or ARCH errors yield completeness. Because completeness, P-completeness (D'Haultfœuille (2011)), and $L^2(f_{X_2^*})$-completeness (Andrews (2011)) each implies $\mathcal{G}$-completeness, sufficient conditions for any of the former are also sufficient for the latter.

While in other contexts completeness conditions may appear to be high-level conditions that are difficult to interpret and verify, I would argue that Assumption ID 4 is reasonable in many economic applications. Intuitively, completeness requires a strong relationship between X_2^* and X_1^* which does not appear problematic in the context of the examples mentioned in the introduction because their latent regressors (determinants of investment decisions, cognitive and noncognitive ability etc.) are highly persistent. Furthermore, Markov models for the latent explanatory variables in those models are

common (see the references in the introduction) and allow for direct use of the low-level sufficient conditions for Assumption ID 4 as discussed in the previous paragraph. Finally, notice that beyond the nonparametric rank condition, the latent transition process is left completely unspecified. In particular, no functional form assumptions are made.

The next set of assumptions called IVX describes the restrictions necessary for X_1 to be a valid instrument for identification of $F_{Y_2|X_2^*}$. Similar assumptions called IVY that can replace IVX to make the outcome variable a valid instrument are given below.

Definition 2. *Let the observable transition operator* $\mathcal{D}_X : L^2(f_{X_2}) \to L^2(f_{X_1})$ *be defined as*

$$\mathcal{D}_X h := E[h(X_2)|X_1 = \cdot], \qquad h \in L^2(f_{X_2}).$$

Assumption IVX 1. *For any fixed* $y_2 \in \mathbb{Y}_2$, *the Fourier transform of* $[\mathcal{D}_X^{-1} F_{Y_2|X_1}(y_2|\cdot)](\cdot)$ *is nonzero on a dense subset of* $\mathbb{R}^p$.

This assumption rules out cases in which the Fourier transform of the quantity $\mathcal{D}_X^{-1} F_{Y_2|X_1}$ vanishes over an interval, but allows for (potentially infinitely many) isolated zeros. $\mathcal{D}_X^{-1} F_{Y_2|X_1}$ being a polynomial seems to be the only non-trivial example violating the assumption, but the proof of the identification result below shows $\mathcal{D}_X^{-1} F_{Y_2|X_1}$ convolved with the ME distribution is bounded, so the polynomial case is unlikely. Appendix A gives sufficient conditions for this assumption. Notice the assumption is stated in terms of observables only and so, in principle, could be tested.

Assumption IVX 2. **(i)** $(Y_2, X_2) \perp (X_1^*, X_1) \mid X_2^*$, **(ii)** $Y_2 \perp X_2 \mid (X_2^*, X_1^*, X_1)$ *and* **(iii)** $X_2^* \perp X_1 \mid X_1^*$.

The first part of Assumption IVX 2 says past observed and latent regressors should not contain any more information about the present outcome and covariate than the present true regressor already does. This condition rules out time dependence in the ME.[7] An extension of these exclusion restrictions to additional, perfectly measured regressors allows dynamic models with past outcomes as additional regressors as well as for feedback from past outcomes to the independent variable. The regression error $Y_t \mid X_t^*$ is allowed to be serially dependent as well as conditionally heteroskedastic: they can depend on contemporaneous regressors but not on past ones. Part (ii) requires the ME to be independent of the structural error $Y_t \mid X_t^*$, a strengthening of the standard uncorrelatedness assumption in linear panel models with ME. The third exclusion restriction is a redundancy condition

---

[7]A second identification argument presented below uses the outcome as an instrument and accommodates time series dependence in the ME.

that assumes past mismeasured covariates do not contain more information about true covariates than past true covariates. In the next subsection, I discuss these conditional independence assumptions in the context of a nonparametric panel data regression model.

With Assumptions ID and IVX, stating the first of two identification results is possible.[8]

Theorem 1. *Under Assumptions ID and IVX, the conditional cdf of Y_2 given X_2^*, $F_{Y_2|X_2^*}$, is identified. If, in addition, the distribution of the ME η_t is the same for all t then the distribution of η, F_η, the transition law $F_{X_2^*|X_1^*}$, and the distribution of X^*, F_{X^*}, are identified as well.*

The theorem shows the structural relationship between the outcome and the latent independent variable as characterized by the conditional cdf $F_{Y_2|X_2^*}$ is identified. In some sense, this distribution embodies all characteristics of the structural model. In discrete choice models, for example, it determines choice probabilities and may be of direct interest. In other settings, the distribution may not be relevant by itself, but I subsequently show how regression functions and marginal effects can be computed from this conditional cdf.

Remark 1. *Part (i) of Theorem 1 identifies only the structural relationship in the second time period. Obviously, if the relationship is stationary over time, the argument identifies it for all time periods. If, on the other hand, $F_{Y_t|X_t^*}$ varies with t, the argument requires T time periods to identify $T-1$ structural relationships.*

Remark 2. *As can be seen in the subsequent sketch of the proof, the identification argument uses only X_1 from the past period but does not involve Y_1. Therefore, Theorem 1 also presents a constructive identification result for cross-sectional models of the following form. Let Y denote an individual's outcome, X^* a latent covariate, and suppose one observes X and Z with*

$$X = X^* + \eta_X, \qquad X^* \perp \eta_X,$$
$$Z = Z^* + \eta_Z, \qquad Z^* \perp \eta_Z.$$

Here Z is a noisy observation of an (unobserved) instrumental variable Z^*, which is assumed to depend on X^*. Identification of such a model appears to be a new result

---

[8]Intuitively, identification of a function here means it possesses a unique representation in terms of observed variables or quantities that can be constructed from observables. For more formal definitions of identification, see Koopmans and Reiersol (1950), Roehrig (1988), and the handbook chapters by Matzkin (1994, 2007).

relative to the existing constructive identification arguments. The latter either require asymmetry between how X and Z relate to X^*,

$$X = X^* + \eta_X, \qquad X^* \perp \eta_X,$$
$$X^* = Z + \eta_Z, \qquad Z \perp \eta_Z,$$
(2)

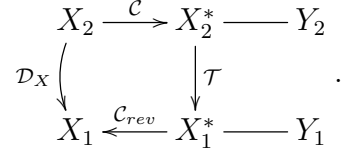(Schennach (2007)), or that X and Z are repeated measurements of the exact same variable,

$$X = X^* + \eta_X, \qquad X^* \perp \eta_X,$$
$$Z = X^* + \eta_Z, \qquad X^* \perp \eta_Z,$$
(3)

(Hausman, Newey, Ichimura, and Powell (1991), Li and Vuong (1998), Schennach (2004a, 2008), and Cunha, Heckman, and Schennach (2010)). Model (2) requires the instrument Z to predict X^* in the sense that X^* fluctuates around the value of Z. For example, Z may be an aggregate measure of the variable X^*. If the second measurement Z is taken at a different point in time or in a different environment than X, one may question whether they really measure the same underlying quantity X^*. The approach of this paper allows for such changes and accommodates instruments Z, which measure a variable related to the variable of interest, but not necessarily the same. Consequently, the requirements an instrument must satisfy are weaker than in existing approaches based on (2) and (3).

Remark 3. *The same argument as in the proof of Theorem 1 can also be used to directly identify $E[Y_t|X_t^* = \cdot]$, provided the conditional expectation function is bounded by polynomials. Unlike for $F_{Y_t|X_t^*}(y|\cdot)$, however, the argument would require dealing with generalized functions so that nonparametric estimation and inference becomes considerably more difficult.*

Remark 4. *In the recent econometrics literature, Kotlarski's lemma (Kotlarski (1967)) has gained some popularity for identifying nonparametric panel data as well as ME models. In a panel data model of the form $Y_t = m(X_t, \alpha) + \varepsilon_t$, but without ME, Evdokimov (2010) views the outcomes in two different periods as repeated measurements of $m(X_t, \alpha)$ given the regressor X_t does not change over those two periods. This approach cannot be expected to work in the present setup because conditioning on the observed regressors to be equal over time does not imply the latent regressors are constant over time as well. Kotlarski's lemma identifies cross-sectional models with ME and repeated measurements $X_1 = X^* + \eta_1$ and $X_2 = X^* + \eta_2$, say, because they assume X_1 and X_2 are measurements of the same underlying latent variable X^*; see the references in Remark 2. This approach is applicable to panel data models only in the special case when X_t^* follows a random walk. Theorem 1, on the other hand, allows for general nonparametric transition processes.*

11

The formal proof of Theorem 1 is given in Appendix B.1. The following discussion provides an overview of the argument. For simplicity, consider the case of univariate covariates X_t^*. First, I introduce the following short-hand notation: the contamination operator $\mathcal{C}h := E[h(X_2)|X_2^* = \cdot]$, the reverse contamination operator $\mathcal{C}_{rev}h := E[h(X_1^*)|X_1 = \cdot]$, and the latent transition operator $\mathcal{T}h := E[h(X_2^*)|X_1^* = \cdot]$. The following graph illustrates the relationship between the random variables in the model and the operators just introduced:

$$
\begin{array}{ccccc}
X_2 & \xrightarrow{\mathcal{C}} & X_2^* & \!\!\!\!—\!\!\!\! & Y_2 \\
{\scriptstyle \mathcal{D}_X}\downarrow & & \downarrow{\scriptstyle \mathcal{T}} & & \\
X_1 & \xleftarrow{\mathcal{C}_{rev}} & X_1^* & \!\!\!\!—\!\!\!\! & Y_1
\end{array}
.
$$

For example, in the case of $\mathcal{C}$, functions of X_2 are mapped to functions of X_2^*. With this notation at hand and under appropriate conditional independence assumptions, one can write the observable quantities $d_Y(x_1) := F_{Y_2|X_1}(y_2|x_1)$ and $d_{YX}(x_1) := E[X_2 \mathbb{1}\{Y_2 \le y_2\}|X_1]$ as

$$
d_Y(x_1) = \iint F_{Y_2|X_2^*}(y_2|x_2^*) f_{X_2^*|X_1^*}(x_2^*|x_1^*) f_{X_1^*|X_1}(x_1^*|x_1) dx_2^* dx_1^*
$$

and

$$
d_{YX}(x_1) = \iint x_2^* F_{Y_2|X_2^*}(y_2|x_2^*) f_{X_2^*|X_1^*}(x_2^*|x_1^*) f_{X_1^*|X_1}(x_1^*|x_1) dx_2^* dx_1^*.
$$

With $\tilde{F}_{Y_2|X_2^*}(y_2|x_2^*) := x_2^* F_{Y_2|X_2^*}(y_2|x_2^*)$ and the operator notation just introduced, these two equations can be rewritten as

$$
d_Y = \mathcal{C}_{rev}\mathcal{T}F_{Y_2|X_2^*}, \tag{4}
$$

$$
d_{YX} = \mathcal{C}_{rev}\mathcal{T}\tilde{F}_{Y_2|X_2^*}, \tag{5}
$$

where the operator $\mathcal{C}_{rev}\mathcal{T}$ is applied with respect to x_2^*, keeping y_2 as a fixed parameter. Next, the figure above suggests the observable transition in the covariates (represented by $\mathcal{D}_X$) consists of three intermediate transitions: (i) reversing the contamination from X_1 to X_1^*, (ii) performing the latent transition from X_1^* to X_2^*, and (iii) contaminating X_2^* with ME to get X_2. In terms of operators, we thus have the identity $\mathcal{D}_X = \mathcal{C}_{rev}\mathcal{T}\mathcal{C}$. Solving for $\mathcal{C}_{rev}\mathcal{T}$ and substituting the expression into (4) and (5) yields

$$
\mathcal{C}\mathcal{D}_X^{-1}d_Y = F_{Y_2|X_2^*},
$$

$$
\mathcal{C}\mathcal{D}_X^{-1}d_{YX} = \tilde{F}_{Y_2|X_2^*}.
$$

Rewrite this system of equations in the form of convolution equations

$$\mathcal{C}s_Y = F_{Y_2|X_2^*}, \tag{6}$$

$$\mathcal{C}s_{YX} = \tilde{F}_{Y_2|X_2^*}, \tag{7}$$

in which the functions s_Y and s_{YX} solve the two integral equations

$$d_Y(x_1) = E[s_Y(X_2)|X_1 = x_1], \tag{8}$$

$$d_{YX}(x_1) = E[s_{YX}(X_2)|X_1 = x_1]. \tag{9}$$

Notice (6) and (7) differ from the usual convolution equations encountered in related work: convolving the *observed* functions s_Y and s_{YX} with the distribution of the ME produces the *unobserved* function we want to identify. Typically, the roles of observed and unobserved quantities are reversed (e.g. Schennach (2004a,b, 2007, 2008)).

Since the two equations (8) and (9) involve only observable quantities, s_Y and s_{YX} are identified. The $\mathcal{G}$-completeness condition is required to show these two functions are in fact the unique solutions to (8) and (9), respectively. Finally, taking Fourier transforms of both convolution equations (6) and (7) yields two algebraic equations with two unknown functions, the characteristic function of the ME and the Fourier transform of the conditional cdf. The formal argument is more involved because one cannot simply take Fourier transforms of conditional cdfs (as functions of their conditioning argument). The system can then be solved and results in explicit expressions as summarized in the following corollary.

Corollary 1. *Let $\mathcal{F}$ denote the Fourier transform operator, $\mathbf{i} := \sqrt{-1}$, Γ a smooth path connecting 0 and ζ in $\mathbb{R}^p$, and let $\nabla_z f$ denote the gradient vector of a function f with respect to z. Then, under Assumptions ID and IVX,*

$$F_{Y_2|X_2^*}(y_2|x_2^*) = \frac{1}{2\pi}\int \phi(\zeta)\sigma_Y(\zeta)e^{-\mathbf{i}\zeta x_2^*}d\zeta + c_Y(y_2),$$

where

$$\phi(\zeta) := \exp\left\{\int_0^\zeta \frac{\mathbf{i}\sigma_{YX}(z) - \nabla_z\sigma_Y(z)}{\sigma_Y(z)}\cdot d\Gamma(z)\right\},$$

the Fourier transforms $\sigma_Y(\zeta) := [\mathcal{F}\mathcal{D}_X^{-1}d_Y](\zeta)$ and $\sigma_{YX}(\zeta) := [\mathcal{F}\mathcal{D}_X^{-1}d_{YX}](\zeta)$, the observable conditional expectations $d_Y(x_1) := F_{Y_2|X_1}(y_2|x_1)$ and $d_{YX}(x_1) := E[X_2\mathbb{1}\{Y_2 \leq y_2\}|X_1 = x_1]$, and finally the centering constants

$$c_Y(y_2) := \lim_{R_1\to\infty}\lim_{R_2\to\infty}\frac{\int_{R_1\leq|x_2|\leq R_2} F_{Y_2|X_2}(y_2|x_2)dx_2}{\int_{R_1\leq|x_2|\leq R_2} dx_2}.$$

13

I now turn to a variant of the above identification argument that gives assumptions such that a past outcome can replace the past mismeasured covariate as the instrument.

Definition 3. *For $1 \leq r < s \leq T$, introduce the notation $Y_{r:s} := (Y_r, \ldots, Y_s)'$. Define the observable transition operator $\mathcal{D}_Y : L^2(f_{X_T}) \to L^2(f_{Y_{1:T-1}})$ as*

$$\mathcal{D}_Y h := E[h(X_T)|Y_{1:T-1} = \cdot], \qquad h \in L^2(f_{X_T}).$$

Also, let $\mathcal{M} : L^2(f_{Y_{1:T-1}}) \to L^2(f_{X^*_{T-1}})$ be the latent model operator defined as

$$\mathcal{M} h := E[h(Y_{1:T-1})|X^*_{T-1} = \cdot], \qquad h \in L^2(f_{Y_{1:T-1}}).$$

In the second identification result below, the vector $Y_{1:T-1}$ contains all the past outcomes and serves as the instrument vector for recovering the structural relationship in period T. Because outcomes are scalar and identification requires at least as many instruments as mismeasured covariates, the number of observed time periods must exceed the dimension of the mismeasured covariate by at least one ($T \geq p + 1$).

The following two assumptions replace the similar counterparts, Assumptions IVX 1 and 2.

Assumption IVY 1. *For any fixed $y_T \in \mathbb{Y}_T$, the Fourier transform of the function $[\mathcal{D}_Y^{-1} F_{Y_T|Y_{1:T-1}}(y_T|\cdot)](\cdot)$ is nonzero on a dense subset of $\mathbb{R}^p$.*

Assumption IVY 2. *Suppose* **(i)** $(Y_T, X_T) \perp (X^*_{T-1}, Y_{1:T-1}) \mid X^*_T$, **(ii)** $Y_T \perp X_T \mid (X^*_T, X^*_{T-1}, Y_{1:T-1})$, *and* **(iii)** $X^*_T \perp Y_{1:T-1} \mid X^*_{T-1}$.

Notice Assumption IVY 2(i) requires the structural error $Y_t \mid X^*_t$ to be independent over time. By parts (ii) and (iii), past outcomes must be excluded from the outcome equation and from the latent transition process.

Assumption IVY 3. **(i)** $Y_t = m(X^*_t) + \varepsilon_t$ *with* $\varepsilon_t \perp X^*_t$ *for all* t; **(ii)** $m_2(X^*_t)$ *and* ε_t *have a density with respect to Lebesgue measure on $\mathbb{R}$ and support on whole $\mathbb{R}$;* **(iii)** f_{X^*} *is bounded and there is a constant $C \in \mathbb{R}$ such that $f_{X^*}(x^*) \leq C(1 + |x^*|^2)^{-1}$ for all $x^* \in \mathbb{R}^{Tp}$;* **(iv)** *The characteristic function of ε_t is infinitely often differentiable almost everywhere and nonzero on $\mathbb{R}$.*

The independence assumption and additive separability in part (i) are strong restrictions, representing the cost of unrestricted time series dependence in the ME and of the constructive approach taken in this paper. The assumption restricts the structural relationship in a way that ensures outcomes contain enough information about the latent

14

explanatory variables. Because the latent explanatory variables are continuous random variables with support on $\mathbb{R}^p$, part (ii) ensures that the instruments (the past outcomes) have the same support. The remaining two parts are similar to assumptions made above and satisfied by most standard distributions.

With Assumptions ID and IVY, the next theorem summarizes the identification result based on outcomes as instruments.

Theorem 2. *Suppose $T = p+1$, and Assumptions ID and IVY hold. Then the conditional cdf of Y_T given X_T^*, $F_{Y_T|X_T^*}$, is identified. If, in addition, the distribution of the ME η_t is the same for $t \in \{T-1, T\}$, then the distribution of $\eta_{T-1:T}$, $F_{\eta_{T-1:T}}$, the transition law $F_{X_T^*|X_{T-1}^*}$, and the distribution of $X_{T-1:T}^*$, $F_{X_{T-1:T}^*}$ are identified as well.*

Remarks similar to those stated after Theorem 1 apply here as well. The proof of the theorem is similar in spirit to the one of Theorem 1. One difference, however, is worth pointing out. The discussion relies on adjoint operators, so I have to be explicit about the inner products associated with the various weighted L^2-spaces. For any density f occurring in this paper, equip the space $L^2(f)$ with the usual inner product $\langle h_1, h_2 \rangle := \int h_1(u)h_2(u)f(u)du$ when $h_1, h_2 \in L^2(f)$. With a slight abuse of notation, I denote all inner products and the induced norms in the different L^2-spaces by the same symbols, $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. Which space they refer to should be clear from the context. Consider the univariate case $p = 1$ and, analogously to the conditional expectation operators defined above, let $\mathcal{C}$ now be the contamination operator in the T-th period and $\mathcal{M}^*$ the adjoint of $\mathcal{M}$. The following figure illustrates the relationships between the relevant operators and random variables:

$$
\begin{array}{ccc}
X_2 \xrightarrow{\;\mathcal{C}\;} X_2^* \text{——} Y_2 & & \\
\qquad\;\; \downarrow{\scriptstyle \mathcal{T}} & & . \\
X_1 \text{——} X_1^* \xrightarrow{\;\mathcal{M}^*\;} Y_1 & &
\end{array}
$$

Using the definitions from Corollary 2 and $\tilde{F}_{Y_T|X_T^*}(y_T|x_T^*) := x_T^* F_{Y_T|X_T^*}(y_T|x_T^*)$, consider the two integral equations

$$d_Y^\circ = \mathcal{M}^* \mathcal{T} \, F_{Y_T|X_T^*}, \tag{10}$$

$$d_{YX}^\circ = \mathcal{M}^* \mathcal{T} \, \tilde{F}_{Y_T|X_T^*}, \tag{11}$$

which follow from similar arguments as those that led to equations (4) and (5). Assumption IVY 3 guarantees the observable transition from Y_1 to X_2, represented by

$\mathcal{D}_Y = \mathcal{M}^*\mathcal{T}\mathcal{C}$, is invertible and can be used to substitute out the unobserved transition $\mathcal{T}$ in (10) and (11):

$$\mathcal{C}\mathcal{D}_Y^{-1}d_Y^\circ = F_{Y_T|X_T^*},$$
$$\mathcal{C}\mathcal{D}_Y^{-1}d_{YX}^\circ = \tilde{F}_{Y_T|X_T^*}. \tag{12}$$

With these two equations, the remainder of the proof closely follows that of Theorem 1. The illustrating figures above suggest a certain symmetry between X_t and Y_t in the sense that one can view both as measurements of the underlying latent variable X_t^*. Notice, however, the two identification arguments based on covariates or outcomes as instruments are not symmetric at all. The reason for the asymmetry lies in the different strategies to extract information about the latent transition $\mathcal{T}$ from observable transitions. Even though it is reasonable to assume a simple convolution relationship between observed and latent covariates, $X_t = X_t^* + \eta_t$, we want to allow for structural relationships more general than that. Identification based on past covariates as instruments can use the observed transition from X_1 to X_2 for backing out the unobserved transition, but when the instruments are past outcomes, the observed transition from Y_1 to Y_2 would not lead to a solvable convolution problem such as (12). Instead, the present identification argument is based on the transition from Y_1 to X_2, exploiting the convolution relationship between X_2^* and X_2, and leads to the desired form (12).

Analogously to Corollary 1, the following corollary provides the expression of the conditional cdf $F_{Y_T|X_T^*}(y_T|x_T^*)$ in terms of observables.

Corollary 2. *Let $\mathcal{F}$ denote the Fourier transform operator, $\mathbf{i} := \sqrt{-1}$, Γ a smooth path connecting 0 and ζ in $\mathbb{R}^p$, and let $\nabla_z f$ denote the gradient vector of a function f with respect to z. Then, under the assumptions of Theorem 2,*

$$F_{Y_T|X_T^*}(y_T|x_T^*) = \frac{1}{2\pi}\int \phi(\zeta)\sigma_Y^\circ(\zeta)e^{-\mathbf{i}\zeta x_1^*}d\zeta + c_Y^\circ(y_T),$$

where

$$\phi(\zeta) := \exp\left\{\int_0^\zeta \frac{\mathbf{i}\sigma_{YX}^\circ(z) - \nabla_z\sigma_Y^\circ(z)}{\sigma_Y^\circ(z)} \cdot d\Gamma(z)\right\},$$

the Fourier transforms $\sigma_Y^\circ(\zeta) := [\mathcal{F}\mathcal{D}_Y^{-1}d_Y^\circ](\zeta)$ and $\sigma_{YX}^\circ(\zeta) := [\mathcal{F}\mathcal{D}_Y^{-1}d_{YX}^\circ](\zeta)$, the conditional expectations $d_Y^\circ(y_{1:T-1}) := F_{Y_T|Y_{1:T-1}}(y_T|y_{1:T-1})$ and $d_{YX}^\circ(y_{1:T-1}) := E[X_T\mathbb{1}\{Y_T \leq y_T\}|Y_{1:T-1} = y_{1:T-1}]$, and finally the centering constants

$$c_Y^\circ(y_T) := \lim_{R_1\to\infty}\lim_{R_2\to\infty}\frac{\int_{R_1\leq|x_T|\leq R_2}F_{Y_T|X_T}(y_T|x_T)dx_T}{\int_{R_1\leq|x_T|\leq R_2}dx_T}.$$

16

2.2 Nonparametric Panel Data Regression with Fixed Effects

In this subsection, I apply the instrumental variable identification approach from the previous subsection to a nonparametric panel data regression model with fixed effects and ME. Specifically, consider

$$\tilde{Y}_t = m(\tilde{X}_t^*) + \alpha + \varepsilon_t$$
$$\tilde{X}_t = \tilde{X}_t^* + \tilde{\eta}_t \tag{13}$$

with individual-specific heterogeneity α and $T = 4$ time periods. The dependence between α and $\tilde{X}_t^*$ is left completely unrestricted. Porter (1996) and Henderson, Carroll, and Li (2008), for example, study such a model without ME. Defining $\Delta\tilde{Y}_4 := \tilde{Y}_4 - \tilde{Y}_3$ (and similarly for other variables), and

$$Y_2 := \Delta\tilde{Y}_4 \qquad X_2^* := (\tilde{X}_3^*, \tilde{X}_4^*) \qquad X_2 := (\tilde{X}_3, \tilde{X}_4) \qquad \eta_2 := (\tilde{\eta}_3, \tilde{\eta}_4)$$
$$Y_1 := \Delta\tilde{Y}_2 \qquad X_1^* := (\tilde{X}_1^*, \tilde{X}_2^*) \qquad X_1 := (\tilde{X}_1, \tilde{X}_2) \qquad \eta_1 := (\tilde{\eta}_1, \tilde{\eta}_2)$$

then allows application of Theorem 1 or Theorem 2 to (Y, X, X^*, η), resulting in identification of $F_{\Delta\tilde{Y}_t | \tilde{X}_t^*, \tilde{X}_{t-1}^*}$. In the remainder of this subsection, I discuss lower-level sufficient conditions for the hypotheses of the two theorems, using the particular structure imposed in (13). I also show how knowledge of $F_{\Delta\tilde{Y}_t | \tilde{X}_t^*, \tilde{X}_{t-1}^*}$ identifies m and marginal effects.

For the case with past covariates as outcomes, consider the following two assumptions on the dependence between the variables in the model.

Assumption REG 1. **(i)** $\varepsilon_t \perp \{\tilde{X}_s^*\}_{s<t} \mid \tilde{X}_t^*$ and $\varepsilon_t \perp \{\tilde{X}_s^*\}_{s<t} \mid (\tilde{X}_t^*, \varepsilon_{t-1})$, **(ii)** $\varepsilon \perp \tilde{\eta} \mid \tilde{X}^*$, and **(iii)** $\tilde{\eta}_t \perp \tilde{\eta}_s$ for all $s \neq t$.

Part (i) of this assumption allows for contemporaneous heterogeneity in the regression error ε_t, so it could be of the form $\varepsilon_t = \sigma(\tilde{X}_t^*)u_t$ with u_t independent of all other variables in the model. Serial dependence in the regression error is permitted as well, but part (iii) rules out serial dependence in the ME. The remainder of the assumption requires the ME and the regression error to be independent, just as in the linear model with ME (Griliches and Hausman (1986)).

Assumption REG 2. $E[\varepsilon_t | \tilde{X}_t^*, \tilde{X}_s^*] = 0$ for $s \in \{t-1, t+1\}$.

The following lemma shows these assumptions are sufficient for Assumption IVX 1(i) and (ii), and the conditional mean function m is identified.

Lemma 1. *Under Assumptions ID, 1, 2(iii), and REG 1-2, m is identified up to an additive constant. If, in addition, the distribution of the ME η_t is the same for all t then*

the distribution of $\tilde{\eta}$, $F_{\tilde{\eta}}$, the transition law $F_{X_2^*|X_1^*}$, and the distribution of $\tilde{X}^*$, $F_{\tilde{X}^*}$, are identified as well.

Remark 5. *Just as in the linear panel model with fixed effects, additive time-invariant effects are not identified. They could be identified, of course, under the additional assumption that m passes through some known point or that $E\alpha = 0$.*

Remark 6. *In a semiparametric specification of the form $m(\tilde{x}^*; \theta)$, a finite-dimensional parameter θ is identified directly from $F_{\Delta\tilde{Y}_4|\tilde{X}_4^*,\tilde{X}_3^*}(\Delta y|\tilde{x}_4^*, \tilde{x}_3^*; \theta)$, given standard maximum likelihood assumptions. In this case, it suffices to identify $\sigma_Y(\zeta)$ for only a finite number of values ζ, so one could weaken the requirement that the Fourier transform of $\mathcal{D}_X^{-1}F_{Y_2|X_1}$ is nonzero almost everywhere.*

Remark 7. *One can allow for serial dependence in the ME when there are more than four time periods. As in the linear model (Griliches and Hausman (1986)), regressors sufficiently far in the past are valid instruments as long as the ME is serially independent beyond some finite lag. Specifically, suppose the ME follows a moving average process of order q, $MA(q)$, and that $T \geq 4 + 2q$. Then, for $t \geq 2q + 4$, the regressors $\tilde{X}_{t-2q-2}$ and $\tilde{X}_{t-2q-3}$ can be used as instruments to identify the structural relationship in periods t and $t-1$.*

Remark 8. *Identification in the presence of additional, perfectly measured regressors simply requires conditioning all operations on those variables.*

Because of $E[\varepsilon_t|\tilde{X}_t^*, \tilde{X}_s^*] = 0$, the function m is directly identified from $m(x_2^*) = const. + \int \Delta y\, dF_{\Delta\tilde{Y}_4|\tilde{X}_4^*,\tilde{X}_3^*}(\Delta y|\tilde{x}_2^*, 0)$ with $F_{\Delta\tilde{Y}_4|\tilde{X}_4^*,\tilde{X}_3^*}$ as defined in Corollary 1. Alternatively, assume $Q_{\Delta\varepsilon_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(\tau|\tilde{x}_t^*, \tilde{x}_{t-1}^*)$, the τ-th conditional quantile of the difference in the structural errors given the latent regressors, is equal to zero[9]. Then

$$F_{\Delta\tilde{Y}_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(q + m(\tilde{x}_t^*) - m(\tilde{x}_{t-1}^*)|\tilde{x}_t^*, \tilde{x}_{t-1}^*) = F_{\Delta\varepsilon_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(q|\tilde{x}_t^*, \tilde{x}_{t-1}^*) = \tau$$

if $q = Q_{\Delta\varepsilon_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(\tau|\tilde{x}_t^*, \tilde{x}_{t-1}^*) = 0$. Therefore, the difference in the regression function at different time points is identified by the conditional quantile

$$m(\tilde{x}_t^*) - m(\tilde{x}_{t-1}^*) = Q_{\Delta\tilde{Y}_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(\tau|\tilde{x}_t^*, \tilde{x}_{t-1}^*).$$

Also, under the stronger assumption that ε is independent of the latent regressor, we have

$$m(\tilde{x}_t^*) - m(\tilde{x}_{t-1}^*) = q - F_{\Delta\varepsilon_t}^{-1}\left(F_{\Delta\tilde{Y}_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(q|\tilde{x}_t^*, \tilde{x}_{t-1}^*)\right),$$

---

[9]For instance, the difference between two i.i.d. errors ε_1, ε_2 is symmetric and thus the median of the difference is equal to zero. In general, however, a conditional quantile being zero is difficult to characterize; see also Khan, Ponomareva, and Tamer (2011).

where $F_{\Delta\varepsilon_t}(q)$ is identified by $F_{\Delta\tilde{Y}_t|\tilde{X}_t^*,\tilde{X}_{t-1}^*}(q|c,c)$ for some constant $c \in \mathbb{R}$. The difference $m(\tilde{x}_t^*) - m(\tilde{x}_{t-1}^*)$ directly identifies the effect of a discrete change in the latent regressors on the outcome. Alternatively, marginal changes $\partial m(\tilde{x}_t^*)/\partial\tilde{x}_t^*$ are identified as soon as m itself is identified.

The following result provides conditions under which the model (13) is identified when using past outcomes as instruments.

Assumption REG 3. (i) $\varepsilon_t \perp \{\tilde{X}_s^*\}_{s<t} \mid \tilde{X}_t^*$ and $\varepsilon_t \perp \{\tilde{X}_s^*\}_{s<t} \mid (\tilde{X}_t^*, \varepsilon_{t-1})$, (ii) $\varepsilon \perp \tilde{\eta} \mid \tilde{X}^*$, and (iii) $\varepsilon_t \perp \varepsilon_s \mid \tilde{X}^*$ for all $s \neq t$.

Lemma 2. *Suppose $T = 2(p+1)$ and Assumptions ID, IVY 1, 2(iii), 3, and REG 2 and 3 hold. Then m is identified up to an additive constant. If, in addition, the distribution of the ME η_t is the same for $t \in \{T-1, T\}$ then the distribution of $\tilde{\eta}_{T-1:T}$, $F_{\tilde{\eta}_{T-1:T}}$, the transition law $F_{X_T^*|X_{T-1}^*}$, and the distribution of $\tilde{X}_{T-1:T}^*$, $F_{\tilde{X}_{T-1:T}^*}$, are identified as well.*

Discussion   The comparative advantages of the two identification strategies, using co-variates or outcomes as instruments, lie in the strategies' ability to handle serial dependence in the structural error or in the ME. Temporal dependence in the structural error may be important for a variety of well-known reasons such as omitted variables, omitted individual-specific effects, or misspecified dynamics. On the other hand, in some applications, serial dependence in the ME could be considered an important deviation from the classical ME assumptions. For example, suppose the ME is of the form $\eta_t = \beta + \nu_t$, where β is an individual-specific effect, persistent over time, and ν_t an i.i.d. error. In survey data, β could be interpreted as an individual's ability to answer a question correctly, which may be correlated with other characteristics of the subject, for example, language skills or the person's ability to recall past events. Serial dependence in the ME could also arise from economic theory directly, for instance, as in Erickson and Whited (2000), or due to manager fixed effects as in Bertrand and Schoar (2003). A third possibility occurs when the researcher observes mismeasured flow variables, but the true explanatory variable in the model represents the corresponding stock variable, so the ME in a certain period consists of the sum of past ME's and, thus, exhibits serial correlation by construction. Finally, Bound and Krueger (1991) find serially correlated ME in a validation study of earnings data.

The exclusion restrictions for the different instruments (Assumptions REG 1 and REG 3) are analogous to each other, but the approach based on outcomes as instruments requires an additional restrictions on the structural relationship that are not present when using covariates as instruments. The additional assumption (Assumption IVY 3) imposes distributional assumptions and additive separability on the structural error.

In conclusion, given a particular application, the choice of instrument should be guided by the relative importance of serial dependence in the structural error and the ME, and whether the aforementioned distributional restrictions on the structural error are reasonable.

3 Estimation

The identification arguments in the previous section are constructive and suggest a non-parametric plug-in estimator for regression functions and all unobserved distributions. This section describes the estimation procedure with past covariates as instruments and, to reduce the notational burden, considers only panel data models with univariate regressors. However, estimation based on outcomes as instruments can be carried out analogously, and multivariate extensions are straightforward. Following the introduction of the estimator, I present its uniform convergence rate and conditions for uniform consistency.

3.1 Construction of the Estimator

Suppose we observe an i.i.d. sample $\{(y_{i,1}, x_{i,1}, y_{i,2}, x_{i,2})\}_{i=1}^{n}$ of (Y_1, X_1, Y_2, X_2). I suggest an estimator of $F_{Y_2|X_2^*}(y_2|x_2^*)$ based on the following procedure:

Step 1: Construct regularized series estimators $\hat{s}_Y$ and $\hat{s}_{YX}$ of the solutions to the two ill-posed inverse problems $\hat{\mathcal{D}}_X s_Y = \hat{d}_Y$ and $\hat{\mathcal{D}}_X s_{YX} = \hat{d}_{YX}$, respectively.

Step 2: Take Fourier transforms of $\hat{s}_Y$ and $\hat{s}_{YX}$, resulting in estimators $\hat{\sigma}_Y$ and $\hat{\sigma}_{YX}$.

Step 3: Combine $\hat{\sigma}_Y$ and $\hat{s}_{YX}$ to an estimate $\hat{\phi}$ of ϕ using the formula from the identification argument. Substituting $\hat{\phi}$ and $\hat{\sigma}_Y$ into the expression for $F_{Y_2|X_2^*}$ then yields an estimator $\hat{F}_{Y_2|X_2^*}$.

Estimation of regression models requires an additional step:

Step 4: Compute either the conditional mean or conditional quantile function of $\hat{F}_{Y_2|X_2^*}$ as an estimator of the regression function m.

I proceed by describing the four steps in more detail and formally define the estimators for the subsequent derivation of asymptotic properties.

Step 1: Inverting the Observed Transition Operator   The first step of the estimation procedure requires solving finite sample counterparts of the two ill-posed inverse problems, $\mathcal{D}_X s_Y = d_Y$ and $\mathcal{D}_X s_{YX} = d_{YX}$. Consider the first equation, which, by the definition of the observed transition operator $\mathcal{D}_X$, is equivalent to

$$E[s_Y(X_2)|X_1 = x_1] = d_Y(x_1).$$

Estimating the solution s_Y to this equation poses an inherently difficult statistical problem. It is further complicated by the fact that the density $f_{X_2|X_1}$ and the function d_Y are not known and need to be estimated as well. Theorem 1 guarantees the existence of a unique solution s_Y to the population problem, but it is not a continuous functional of d_Y. The discontinuity is implied by the underlying function space being infinite dimensional. As an important consequence, direct application of $\mathcal{D}_X^{-1}$ may blow up small estimation errors in $\hat{d}_Y$, leading to inconsistent estimates of s_Y.

The following nonparametric estimator of s_Y sufficiently regularizes the problem and facilitates consistent estimation.

Definition 4. *Let $u_i := F(x_{i,1})$ be the transformed realization of the first-period explanatory variable with $F : \mathbb{R} \to [0,1]$ a continuous, strictly increasing function. Let $\lambda_{[0,1]}$ denote Lebesgue measure on $[0,1]$. For bases $\{b_j(u)\}_{j\geq 1}$ and $\{p_j(u)\}_{j\geq 1}$ of $L^2(\lambda_{[0,1]})$ and $L^2(f_{X_2})$, respectively, define $B^{J_n}(u) := (b_1(u), \ldots, b_{J_n}(u))'$ and $\mathbf{B} := (B^{J_n}(u_1), \ldots, B^{J_n}(u_n))$, $P^{K_n}(x_2) := (p_1(x_2), \ldots, p_{K_n}(x_2))'$ and $\mathbf{P} := (P^{K_n}(x_{1,2}), \ldots, P^{K_n}(x_{n,2}))$. For some fixed y_2, let $\Upsilon_Y := (\mathbb{1}\{y_{1,2} \leq y_2\}, \ldots, \mathbb{1}\{y_{n,2} \leq y_2\})'$, $\Upsilon_{YX} := (x_{1,2}\mathbb{1}\{y_{1,2} \leq y_2\}, \ldots, x_{n,2}\mathbb{1}\{y_{n,2} \leq y_2\})'$, let I be the $K_n \times K_n$ identity matrix, and A^- the generalized inverse of a matrix A. Then define the series estimators $\hat{s}_Y$ and $\hat{s}_{YX}$ of s_Y and s_{YX} as*

$$\hat{s}_Y(x_2) := P^{K_n}(x_2)'\hat{\beta}_Y,$$
$$\hat{s}_{YX}(x_2) := P^{K_n}(x_2)'\hat{\beta}_{YX},$$

with the series coefficients

$$\hat{\beta}_Y := \left(\mathbf{P}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{P} + \alpha_n I\right)^- \mathbf{P}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\Upsilon_Y, \tag{14}$$
$$\hat{\beta}_{YX} := \left(\mathbf{P}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{P} + \alpha_n I\right)^- \mathbf{P}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\Upsilon_{YX}. \tag{15}$$

The estimators are similar to the series estimator in Hall and Horowitz (2005) and Blundell, Chen, and Kristensen (2007). It takes the form of the two-stage least-squares estimator of the regression of Υ_Y and Υ_{YX} onto $\mathbf{B}$ using $\mathbf{P}$ as an instrument, except the penalty term $\alpha_n I$ in the denominator[10]. The parameter α_n is required to vanish as the

---

[10]Blundell, Chen, and Kristensen (2007) recommend to add the penalization term $\alpha_n I$ because it may improve the finite sample performance of the resulting estimator. However, for consistency, this term is not necessary because the ill-posed inverse problem is regularized by truncating the bases $\{b_j\}$ and $\{p_j\}$.

sample size grows, leading to no penalization in the limit. In principle, it would be possible to derive the restrictions on s_Y and s_{YX} implied by the constraint that $F_{Y_2|X_2^*} \in \bar{\mathcal{G}}$. As in Blundell, Chen, and Kristensen (2007), those restrictions could then be imposed by changing the penalty term $\alpha_n I$ accordingly, but I do not pursue this approach here because the regularization by $\alpha_n I$ worked well in the simulations. The conditioning variable X_1 is transformed to $U := F(X_1)$ with support equal to $[0, 1]$ to facilitate the use of existing uniform convergence results for series estimators. F could be chosen as some cdf, for instance. Given smoothing parameters α_n, K_n, and J_n, the computation of $\hat{s}_Y$ and $\hat{s}_{YX}$ requires only matrix multiplication and inversion. In most applications with scalar or bivariate mismeasured covariates, the dimensions of matrices to be inverted ($K_n \times K_n$ and $J_n \times J_n$) tend to be small, leading to simple implementation.

Step 2: Computing Fourier Transforms   In this step, I describe how to compute the Fourier transforms $\hat{\sigma}_Y$ and $\hat{\sigma}_{YX}$, estimating σ_Y and σ_{YX}. These estimators are Fourier transforms of the regularized inverses computed in the previous step. The subsequent asymptotic theory requires these estimated functions converge uniformly over $\mathbb{R}$, which I achieve by trimming their tails as follows.

Definition 5. *For some trimming parameter $\bar{x}_n > 0$ and $y_2 \in \mathbb{Y}_2$, define the limits $c_Y^+(y_2) := \lim_{x_2 \to \infty} F_{Y_2|X_2}(y_2|x_2)$, $c_Y^-(y_2) := \lim_{x_2 \to -\infty} F_{Y_2|X_2}(y_2|x_2)$, and $c_Y(y_2) := (c_Y^+(y_2) + c_Y^-(y_2))/2$. Then,*

$$\check{s}_Y(x_2) := \hat{s}_Y(x_2)\mathbb{1}\{|x_2| \leq \bar{x}_n\} - c_Y^-(y_2)\mathbb{1}\{x_2 < -\bar{x}_n\} + c_Y^+(y_2)\mathbb{1}\{x_2 > \bar{x}_n\},$$

$$\check{s}_{YX}(x_2) := \hat{s}_{YX}(x_2)\mathbb{1}\{|x_2| \leq \bar{x}_n\} - x_2 c_Y^-(y_2)\mathbb{1}\{x_2 < -\bar{x}_n\} + x_2 c_Y^+(y_2)\mathbb{1}\{x_2 > \bar{x}_n\},$$

$$\check{s}_{xY}(x_2) := \mathbf{i}x_2\check{s}_Y(x_2).$$

The trimming parameter $\bar{x}_n$ is required to diverge to ∞ with the sample size, leading to no trimming in the limit. On the interval $[-\bar{x}_n, \bar{x}_n]$, the estimators marked by a hacek are equal to the corresponding estimators with a circumflex, but their tails are set to the limits of the estimators with a circumflex. In many economic models, the conditional cdf $F_{Y_2|X_2^*}(y_2|x_2^*)$ is known to converge to zero or one as the conditioning variable diverges to $-\infty$ or ∞, so, by Lemma 4(i), the centering constant becomes $c_Y(y_2) \equiv 1/2$ with $c_Y^+(y_2) \equiv 1$ and $c_Y^-(y_2) \equiv 0$. In the regression example, this case occurs whenever the regression function diverges to $+\infty$ and $-\infty$ as $x_2^* \to \infty$ and $x_2^* \to -\infty$, respectively. Subsequently, I assume c_Y^+ and c_Y^- are known, but the limiting constants are equivalent to the limits of the observed conditional cdf $F_{Y_2|X_2}$, so they could also be estimated.

I define the desired estimators of the Fourier transforms, $\hat{\sigma}_Y$ and $\hat{\sigma}_{YX}$, as the Fourier transforms of $\check{s}_Y$ and $\check{s}_{YX}$. Since multiplication by $\mathbf{i}x_2$ corresponds to differentiation in

the Fourier domain, one can estimate the derivative $\nabla_\zeta \sigma_Y$ by the Fourier transform of $\check{s}_{xY}$.

Definition 6. *For any $\zeta \in \mathbb{R}$, define the Fourier transforms $\hat{\sigma}_Y(\zeta) := \int \check{s}_Y(x_2) e^{\mathbf{i}\zeta x_2} dx_2$, $\hat{\sigma}_{YX}(\zeta) := \int \check{s}_{YX}(x_2) e^{\mathbf{i}\zeta x_2} dx_2$, and $\hat{\sigma}_{xY}(\zeta) := \int \check{s}_{xY}(x_2) e^{\mathbf{i}\zeta x_2} dx_2$.*

Most modern statistical software packages provide an implementation of the Fast Fourier Transform algorithm, which can perform this step efficiently. In practice, one would compute $\check{s}_Y(x_2)$, $\check{s}_{YX}(x_2)$, and $\check{s}_{xY}(x_2)$ for, say, $K := 2^{\bar{k}}$, $\bar{k} \in \mathbb{N}$, values of x_2 in the convex hull of the data $\{x_{1,2}, \dots, x_{n,2}\}$. After stacking those values for each of the estimators, the resulting vectors can be fed into the Fast Fourier Transform algorithm, yielding output vectors of the same length K. These output vectors are the estimates $\hat{\sigma}_Y(\zeta)$, $\hat{\sigma}_{YX}(\zeta)$, and $\hat{\sigma}_{xY}(\zeta)$ at K corresponding points ζ in the frequency domain.[11]

Step 3: Inverting Fourier Transforms The third step involves estimating the characteristic function of the ME and taking inverse Fourier transforms to get an estimator of the desired conditional cdf $F_{Y_t|X_t^*}$.

Definition 7. *For $x_2^* \in \mathbb{R}$, $y_2 \in \mathbb{Y}_2$, and a trimming parameter $\bar{\zeta}_n > 0$, define*

$$\check{F}_{Y_2|X_2^*}(y_2|x_2^*) := \max \left\{ \min \left\{ \hat{F}_{Y_2|X_2^*}(y_2|x_2^*), 1 \right\}, 0 \right\}$$

with

$$\hat{F}_{Y_2|X_2^*}(y_2|x_2^*) = \frac{1}{2\pi} \int_{|\zeta| \le \bar{\zeta}_n} \hat{\phi}(\zeta) \hat{\sigma}_Y(\zeta) e^{-\mathbf{i}\zeta \cdot x_2^*} d\zeta + c_Y(y_2) \tag{16}$$

and

$$\hat{\phi}(\zeta) := \exp \left\{ \int_0^\zeta \frac{\mathbf{i}\hat{\sigma}_{YX}(z) - \hat{\sigma}_{xY}(z)}{\hat{\sigma}_Y(z)} dz \right\}. \tag{17}$$

Consistent estimation of $F_{Y_2|X_2^*}$ requires $\bar{\zeta}_n \to \infty$ as the sample size grows, leading to no trimming in the limit. This additional trimming in (16) is common in deconvolution problems and necessary because the tails of Fourier transforms are difficult to estimate and need to be cut off to gain uniform consistency. Since, in finite samples, $\hat{F}_{Y_2|X_2^*}(y_2|x_2^*)$ can take values outside of $[0, 1]$, I define the version $\check{F}_{Y_2|X_2^*}(y_2|x_2^*)$, which is constraint to the unit interval.

---

[11]Due to the many different conventions for computing Fourier transforms, one must pay close attention to the requirements of a particular implementation of the Fast Fourier Transform algorithm. For example, can standardize a discrete Fourier transform in a variety of ways, and some algorithms require the input function be sampled only at locations on the positive real line.

Step 4: Computing Regression Functions In regression models, a fourth step is required to compute the regression function from $\check{F}_{Y_2|X_2^*}(y_2|\cdot)$. As discussed in section 2.2, either the conditional mean or the conditional quantile of the regression error being zero facilitates the estimation of the regression function from the conditional expectation or conditional quantile of the outcome variable.

Definition 8. *For some value $\tilde{x}_2^* \in \mathbb{R}$, $y_2 \in \mathbb{Y}_2$, and $\tau \in (0,1)$, define the estimator of the regression function $m(\tilde{x}_2^*)$ as either*

$$\hat{m}(\tilde{x}_2^*) := \int_{\mathbb{Y}_2} y \, d\check{F}_{Y_2|X_2^*}(y|\tilde{x}_2^*) \tag{18}$$

or

$$\hat{m}(\tilde{x}_2^*) := \min_{y_2 \in \mathbb{Y}_2} \left\{ \check{F}_{Y_2|X_2^*}(y_2|\tilde{x}_2^*) \geq \tau \right\}. \tag{19}$$

In finite samples, the integration in (18) has to be performed numerically. One possibility, given a fixed value for $\tilde{x}_2^*$, is to sample random variables from $\check{F}_{Y_2|X_2^*}(y|\tilde{x}_2^*)$ and then compute their mean. Alternatively, integration by parts leads to the formula $\int_{\underline{y}}^{\overline{y}} y \, d\check{F}_{Y_2|X_2^*}(y|\tilde{x}_2^*) = \overline{y} - \int_{\underline{y}}^{\overline{y}} \check{F}_{Y_2|X_2^*}(y|\tilde{x}_2^*) dy$, which is valid as long as $\underline{y}$ and $\overline{y}$ are finite. In practice, one can select these lower and upper integration bounds as the minimum and maximum values of outcomes observed in the sample and then perform a standard numerical integration step to compute $\int_{\underline{y}}^{\overline{y}} \check{F}_{Y_2|X_2^*}(y|\tilde{x}_2^*) dy$. Both approaches approximate the quantity in (18) but only incur a numerical error, which can be kept as small as desired, independently of the sample size.

For a given $\tilde{x}_2^*$, the quantile in (19) is estimated by computing $\check{F}_{Y_2|X_2^*}(y_2|\tilde{x}_2^*)$ over a grid of y_2 values, ordering the estimates, and keeping the smallest value that is larger than τ.

Remark 9. *Notice the two estimators of m do not depend on the particular values for y_2 and τ. While deriving optimal choices is beyond the scope of this paper, some recommendations can be given. Since extremal quantiles and the tails of cdf's are difficult to estimate, y_2 and τ should neither be too small nor too large, but perhaps somewhere near the center of the unconditional distribution of Y_2. Also, one may want to compute the estimator $\hat{m}$ for various values of y_2 or τ and then take the average.*

Remark 10. *If m or $F_{Y_2|X_2^*}(y_2|\cdot)$ is known to be monotone, one can estimate either of them for various values $\tilde{x}_2^*$ and then simply sort the estimates in ascending or descending order. The resulting estimator performs well in finite samples as shown in Chernozhukov, Fernández-Val, and Galichon (2010). In addition, imposing monotonicity directly on the estimators of s_Y and s_{YX} may significantly improve their finite sample properties*

24

(Chetverikov and Wilhelm (2015)). In simple examples such as joint normal distribution of (Y, η, X^*) and linearity of m, one can easily find regularity conditions under which s_Y is monotone, but in general it appears difficult to establish whether s_Y satisfies shape restrictions or not. In practice, one may want to use Chetverikov and Wilhelm (2015)'s test for monotonicity of s_Y and impose it if the test fails to reject.

3.2   Uniform Convergence Rates

In this section, I derive uniform convergence rates of the estimators $\check{F}_{Y_2|X_2^*}$ and $\hat{m}$. These rates differ depending on the tail behavior and smoothness of various model components. As before, I focus on the univariate case.

Assumption C 1. *Let $\{(y_{i,1}, x_{i,1}, y_{i,2}, x_{i,2})\}_{i=1}^n$ be an i.i.d. sample of (Y_1, X_1, Y_2, X_2) with $p = 1$. The transformation $F : \mathbb{R} \to [0,1]$ in Definition 4 is such that $U = F(X_1)$ possesses a density that is bounded away from zero over its support $[0,1]$.*

Choosing the function F approximately equal to the empirical cdf of X_1 leads to a transformed random variable U that is close to uniformly distributed over $[0,1]$ and therefore possesses a density bounded away from zero.

Deriving explicit convergence rates for nonparametric estimators typically requires assuming certain quantities involved, such as densities, regression functions, or characteristic functions, belong to some regularity space with known tail behavior or smoothness. First, consider the ill-posed inverse problems of solving

$$E[s_Y(X_2)|X_1 = x_1] = d_Y(x_1) \qquad \text{and} \qquad E[s_{YX}(X_2)|X_1 = x_1] = d_{YX}(x_1) \qquad (20)$$

for s_Y and s_{YX}.

Definition 9. *Denote by $D^d f$ the d-th elementwise derivative of a scalar- or vector-valued function f. Let $\mathcal{H}_n$ be the space of functions $h \in L^2(f_{X_2})$ such that there exists $\Pi \in \mathbb{R}^{K_n}$ with $h = P^{K_n\prime}\Pi$ and $\sum_{k=1}^s \|D^k h\|^2 < c$ for some finite $c \in \mathbb{R}$, $s \in \mathbb{N}$. Then define*

$$\tau_n := \sup_{h \in \mathcal{H}_n: \, h \neq 0} \frac{\|h\|}{\|(\mathcal{D}_X^* \mathcal{D}_X)^{1/2} h\|}.$$

The quantity τ_n is the sieve measure of ill-posedness introduced by Blundell, Chen, and Kristensen (2007). For a given sequence of approximating spaces $\{\mathcal{H}_n\}$, also called sieve, it measures how severely ill-posed the equations in (20) are: a polynomial or exponential divergence rate of τ_n classifies the problem as mildly or severely ill-posed, respectively.

The intuition for this distinction is analogous to the finite-dimensional case in which applying the operator $\mathcal{D}_X$ becomes multiplication by a matrix. Inverting this matrix is difficult when its eigenvalues are close to zero. In the infinite-dimensional case, however, the operator $\mathcal{D}_X$ possesses an infinite number of eigenvalues whose closeness to zero is measured in terms of how fast their ordered sequence converges to zero. The fast (exponential) rate occurs, for example, when X_1 and X_2 are jointly Gaussian, leading to severe ill-posedness.

Assumption C 2. (a) *The smallest eigenvalues of $E[B^{J_n}(F(X_1))B^{J_n}(F(X_1))']$ and $E[P^{K_n}(X_2)P^{K_n}(X_2)']$, respectively, are bounded away from zero uniformly in J_n, K_n;* **(b)** *there is a sequence $\{\omega_{0,n}\}$ in $\mathbb{R}$ such that $\sup_{u \in [0,1]} |B^{J_n}(u)| \leq \omega_{0,n}$;* **(c)** *$p_j(x)$ are bounded uniformly over j and have $\rho_s > 1/2$ square-integrable derivatives;* **(d)** *for any functions $h_p \in L^2(f_{X_2})$ and $h_b \in L^2(\lambda_{[0,1]})$ with l square-integrable derivatives there are Π_p and Π_b such that $\|h_p - P^{K_n'}\Pi_p\| = O(K_n^{-l})$ and $\|h_b - B^{J_n'}\Pi_b\| = O(J_n^{-l})$ as $J_n, K_n \to \infty$;* **(e)** *d_Y, d_{YX} and $E[h(X_2)|X_1 = \cdot]$ have $\rho_d > 1/2$ square-integrable derivatives for all $h \in \mathcal{H}_n$, and each of s_Y, s_{YX}, and s_{xY} has at least $\rho_s \geq 2$ derivatives;* **(f)** *for $k \in \{Y, YX\}$, there is a function $h_k \in \mathcal{H}_n$ so that $\tau_n^2 \|\mathcal{D}_X(s_k - h_k)\|^2 \leq const \cdot \|s_k - h_k\|^2$.*

The various parts of this assumption are standard in the literature on series estimation; see Newey (1997), for example. Part (a) bounds the second moment matrix of the approximating functions away from singularity. Part (b) bounds the individual series terms, which, by the compact support of $U = F(X_1)$, is not restrictive. The third condition assumes the uniform approximation of the target function by a truncated series expansion incurs an error that vanishes at rate $O(K_n^{-\rho})$. Assumption C 2(f) is taken from Blundell, Chen, and Kristensen (2007) and requires that, for some $h_k \in \mathcal{H}_n$, $\mathcal{D}_X h_k$ approximates $\mathcal{D}_X s_k$ at least as well as h_k approximates s_k (after standardizing by τ_n^2). The latter approximation incurs an error that vanishes at the rate given in part (d).

Assumption C 3. *For some $a > \rho_s$, where ρ_s is defined in Assumption C 2(e), $E[|\eta_2|^{2a}] < \infty$ and $E[|X_2^*|^{2a}|X_1^* = x_1^*] < \infty \ \forall x_1^* \in \mathbb{R}$.*

Definition 10. *Given the sequences $\bar{x}_n \to \infty$ and $\bar{\zeta}_n \to \infty$, define the bounds $\underline{\sigma}_n := \inf_{y_2 \in \mathbb{Y}_2} \inf_{|\zeta| \leq \bar{\zeta}_n} |\gamma(\zeta, y_2)/\phi(\zeta)|$ and $\bar{r}_n := \sup_{|\zeta| \leq \bar{\zeta}_n} |\partial \log \phi(\zeta)/\partial \zeta|$. Let*

$$T_n := (\bar{r}_n T_{Y,n} + T_{\Delta,n})\underline{\sigma}_n^{-1}\bar{\zeta}_n^2 + T_{\gamma,n}$$

be a trail-trimming bound with

$$T_{\Delta,n} := \max_{k=0,2} \max_{j=0,1} \max_{d=1,2}\{T_{Y,n}^{d,j,k}, T_{YX,n}^{d,j=0,k}\},$$

$$T_{Y,n} := \max_{d=0,1} \max_{k=0,1} T_{Y,n}^{d,j=0,k},$$

26

and for $L \in \{Y, YX\}$, $d, j, k \in \{0, 1, 2\}$,

$$T_{L,n}^{d,j,k} := \sup_{y_2 \in \mathbb{Y}_2} \bar{\zeta}_n^k \int_{|x_2| > \bar{x}_n} |x_2|^j \left| \nabla^d \hat{s}_L(x_2, y_2) - \nabla^d s_L(x_2, y_2) \right| dx_2,$$

$$T_{\gamma,n} := \sup_{y_2 \in \mathbb{Y}_2} \int_{|\zeta| > \bar{\zeta}_n} |\gamma(\zeta, y_2)| d\zeta.$$

Further, define a bound on the density f_{X_2} by $\underline{f}_n := \inf_{|x_2| \leq \bar{x}_n} f_{X_2}(x_2)$, $\delta_n := K_n^{-\rho_s} + \tau_n \sqrt{K_n/n}$, and $\omega_{d,n} := \sup_{u \in [0,1]} |D^d B^{J_n}(u)|$.

Assuming the existence of upper and lower bounds of Fourier transforms and densities is standard in the literature on deconvolution (e.g., Fan (1991), Fan and Truong (1993), Li and Vuong (1998), and Schennach (2004b)) and nonparametric regression (e.g., Andrews (1995)). The bound $\omega_{d,n}$ depends only on the particular basis chosen. For splines, $\omega_{d,n} = J_n^{1/2+d}$, whereas for orthogonal polynomials, $\omega_{d,n} = J_n^{1+2d}$.

Assumption C 4. *As $n \to \infty$, let the parameter sequences $K_n \to \infty$, $J_n \to \infty$, $\bar{x}_n \to \infty$, $\bar{\zeta}_n \to \infty$, $\alpha_n \to 0$ satisfy $J_n/n \to 0$, $\lim_{n\to\infty}(J_n/K_n) = c > 1$, $\omega_{0,n}^2 K_n/n \to 0$, $\bar{r}_n \delta_n \omega_{1,n}/(\underline{f}_n^{1/2} \bar{\zeta}_n \underline{\sigma}_n) \to 0$, and $\bar{r}_n T_{Y,n}/\underline{\sigma}_n \to 0$.*

Assumption C 4 states rate conditions on the various trimming and smoothing sequences of the nonparametric estimator. Corollaries 3 and 4 below imply these conditions are mutually compatible.

With Assumption C at hand, stating the general form of the uniform convergence rate in terms of the parameter sequences just defined is possible.

Theorem 3. *Let $\bar{\mathbb{X}}$ be some compact subset of $\mathbb{R}$. Under Assumptions ID, IDX, and C,*

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right| = O_p\left(\frac{\omega_{2,n}\bar{x}_n + \bar{r}_n \omega_{1,n}\bar{\zeta}_n}{\underline{f}_n^{1/2} \underline{\sigma}_n} \delta_n + T_n \right).$$

This expression of the convergence rate provides useful information about when the estimator performs well. First of all, $\delta_n := K_n^{-\rho_s} + \tau_n \sqrt{K_n/n}$ is the well-known convergence rate of nonparametric instrumental variable estimators arising from the estimation of s_Y and s_{YX}. Chen and Reiß (2011) and Johannes, Van Bellegem, and Vanhems (2011) provide conditions under which this rate is minimax optimal in the nonparametric instrumental variable context. Second, due to the division by $\underline{f}_n^{1/2}$, the estimator converges slowly if the density of X_2 has thin tails. Intuitively, more variation in X_2 improves the precision of the estimator. Third, the formula for the characteristic function of the ME, equation (17), suggests estimation errors may be large when the Fourier transform $\hat{\sigma}_Y(\zeta)$

in the denominator is close to zero. Because $\sigma_Y(\zeta) = \gamma(\zeta)/\phi(\zeta)$, the Fourier transform $\sigma_Y(\zeta)$ is close to zero whenever $\gamma(\zeta)$, the Fourier transform of $F_{Y_2|X_2^*}(y_2|\cdot)$, is small relative to $\phi(\zeta)$, the characteristic function of the ME. Therefore, thicker tails in the Fourier transform of $F_{Y_2|X_2^*}(y_2|\cdot)$ and a ME characteristic function with thin tails result in a faster convergence rate. Theorem 3 reflects this fact in the division by $\underline{\sigma}_n$. On the other hand, if the characteristic function of the ME decays too quickly, the ill-posed inverse problems may be severely ill-posed, leading to slow convergence rates δ_n. In conclusion, the new estimator is expected to perform well, if (i) the ME's characteristic function has moderately thin tails, (ii) the Fourier transform of the conditional cdf $F_{Y_2|X_2^*}$ has thick tails, and (iii) the density of X_2 has thick tails.

The smoothness of a function determines how thick the tails of its Fourier transform are. The more derivatives the function possesses, the faster the tails of its Fourier transform vanish. To derive explicit rates of convergence in terms of the sample size, the literature on nonparametric estimation typically categorizes functions into different smoothness classes.

Definition 11. *For a function $f(x)$ and a label f, let the expression $f(x) \geq (\leq) \mathcal{S}(f, x)$ mean there exist constants $C_f, \gamma_f \in \mathbb{R}$, and $\alpha_f, \beta_f \geq 0$ such that $\gamma_f \beta_f \geq 0$ and $f(x) \geq (\leq) C_f (1 + |x|)^{\gamma_f} \exp\{-\alpha_f |x|^{\beta_f}\}$ for all x in the domain of f.*

This definition conveniently groups the two common smoothness classes, ordinary smooth and super-smooth functions, into one expression, simplifying the exposition of the explicit rates below. If, for instance, $\beta_f = \alpha_f = 0$, the function f is ordinary smooth; otherwise it is super-smooth.[12]

Assumption S. *Suppose $|\partial \log \phi(\zeta)/\partial \zeta| \leq \mathcal{S}(r, \zeta)$, $\sup_{y_2 \in \mathbb{Y}_2} |\gamma(\zeta, y_2)/\phi(\zeta)| \leq \mathcal{S}(g, \zeta)$, $f_{X_2}(x_2) \geq \mathcal{S}(f_X, x_2)$, and $\sup_{y_2 \in \mathbb{Y}_2} |\partial s_Y(x_2, y_2)/\partial x_2| \geq \mathcal{S}(s, x_2)$.*

This assumption assigns smoothness parameters α, β, and γ to the various quantities whose smoothness is to be classified. Different combinations of values for these parameters generate different convergence rates in terms of the sample size, as summarized in the following two corollaries.

Corollary 3 (Mildly Ill-posed Case). *Let $\bar{\mathbb{X}}$ be some compact subset of $\mathbb{R}$. Suppose $B^{(J_n)}$ is a spline basis, $\tau_n = O(K_n^\omega)$, $\alpha_{f_X} = \beta_{f_X} = 0$ and $\alpha_r = \beta_r = 0$. Let $K_n = O(n^{1/[2(\rho_s+\omega)+1]})$. Then, under Assumptions ID, IDX, C, and S, we have*

---

[12]See Fan (1991) for formal definitions.

(i) $\beta_g > 0$ and $\beta_s > 0$. Suppose $\gamma_r - \beta_g + 3 < 0$ and set $\bar{\zeta}_n = O((\log n)^{1/\beta_g})$. If $\alpha_s > \alpha_g$ and $\rho_s/[2(\rho_s + \omega) + 1] > \alpha_g$, set $\bar{x}_n = O((\log n)^{1/\beta_s})$. If $\bar{\kappa} := 2\alpha_g/\gamma_{f_X} - 2\rho_s/[2\gamma_{f_X}(\rho_s + \omega) + \gamma_{f_X}] > 0$, set $\bar{x}_n = O(n^\kappa)$ with $0 < \kappa < \bar{\kappa}$. In either case,

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right| = O_p\left((\log n)^{(\gamma_r - \beta_g + 3)/\beta_g}\right) = o_p(1).$$

(ii) $\beta_g > 0$ and $\beta_s = 0$. Suppose $\gamma_r - \beta_g + 3 < 0$ and $\gamma_s < -3$. Let $\bar{\zeta}_n = O((\log n)^{1/\beta_g})$ and $\bar{x}_n = O(n^\kappa)$ with $\kappa > -\alpha_g/(\gamma_s + 3)$. Then the convergence rate is the same as in (i).

(iii) $\beta_g = 0$ and $\beta_s > 0$. Suppose $\gamma_r < -3$, let $\bar{\zeta}_n = O(n^{-\underline{m}/(\gamma_g + 1)})$ and $\bar{x}_n = O((\log n)^{1/\beta_s})$ with $\underline{m} := \min\{\alpha_s, \rho_s/[2(\rho_s + \omega) + 1]\}$. Then

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right| = O_p\left(\bar{\beta}(n) n^{-\underline{m}(\gamma_r + 3)/(\gamma_g + 1)}\right) = o_p(1)$$

with

$$\bar{\beta}(n) := \begin{cases} (\log n)^{-\gamma_{f_X}/(2\beta_s)}, & \alpha_s \geq \rho_s/[2(\rho_s + \omega) + 1] \\ (\log n)^{(\gamma_s - 2\beta_s + 3)/\beta_s}, & o.w. \end{cases}.$$

(iv) $\beta_g = 0$ and $\beta_s = 0$. Suppose $\gamma_r, \gamma_s < -3$ and let $\bar{\zeta}_n = O(n^\varpi)$ and $\bar{x}_n = O(n^\varsigma)$ with $\varsigma := \rho_s/[(2(\rho_s + \omega) + 1)(-\gamma_{f_X}/2 - (\gamma_s + 3))]$ and $\varpi := \varsigma(\gamma_s + 3)/(1 + \gamma_g)$. Then

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right| = O_p\left(n^{\varpi(\gamma_r + 3)}\right) = o_p(1).$$

Corollary 4 (Severely Ill-posed Case). *Let $\bar{\mathbb{X}}$ be some compact subset of $\mathbb{R}$. Suppose $B^{(J_n)}$ is a spline basis, $\tau_n = O(\exp\{K_n\})$, $\rho_d = \infty$, and $\alpha_{f_X} = \beta_{f_X} = 0$. Let $K_n = O(\log n)$. Then under Assumptions ID, IDX, C, and S, we have*

(i) $\beta_g > 0$ and $\beta_s > 0$. Set $\bar{\zeta}_n = O((\log \log n)^{1/\beta_r})$ and $\bar{x}_n = O((\log \log n)^{1/\beta_{f_X}})$. Then

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right|$$
$$= O_p\left(\bar{\beta}_1(n)(\log n)^{-\alpha_r + \alpha_{f_X}/2 - \rho_s} + \bar{\beta}_2(n)(\log n)^{-\alpha_r - \alpha_s} + \bar{\beta}_3(n)(\log n)^{-\alpha_r}\right) = o_p(1)$$

with

$$\bar{\beta}_1(n) := (\log \log n)^{(2 + \gamma_r - \gamma_g)/\beta_r - \gamma_{f_X}/(2\beta_{f_X})} \exp\{\alpha_g(\log \log n)^{\beta_g/\beta_r}\},$$
$$\bar{\beta}_2(n) := (\log \log n)^{(2 + \gamma_r - \gamma_g)/\beta_r + (\gamma_s - 2\beta_s + 3)/\beta f_X} \exp\{\alpha_g(\log \log n)^{\beta_g/\beta_r}\},$$
$$\bar{\beta}_3(n) := (\log \log n)^{(\gamma_r - \beta_g + 3)/\beta_r}.$$

29

(ii) $\beta_g > 0$ and $\beta_s = 0$. Set $\bar{\zeta}_n = O((\log \log n)^{1/\beta_r})$ and $\bar{x}_n = O((\log \log n)^{1/\beta_{f_X}})$. Then the convergence rate is the same as in (i) except $\beta_s = \alpha_s = 0$, and it is $o_p(1)$ if $\alpha_r > \alpha_g$, $\beta_r > 0$.

(iii) $\beta_g = 0$ and $\beta_s > 0$. Suppose $\gamma_r < -3$, $\beta_s \geq \beta_{f_X}$ and $\alpha_{f_X}/2 + \varpi(2 + \gamma_r - \gamma_g) - \rho_s < 0$ with $\varpi := -\alpha_s/(1 + \gamma_g)$. Let $\bar{\zeta}_n = O((\log n)^\varpi)$ and $\bar{x}_n = O((\log \log n)^{1/\beta_s})$. Then we have

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right|$$

$$= O_p \left(\bar{\beta}_1(n)(\log n)^{\varpi(2+\gamma_r-\gamma_g)-\rho_s} + \bar{\beta}_2(n)(\log n)^{\varpi(\gamma_r+3)} \right) = o_p(1)$$

with

$$\bar{\beta}_1(n) := (\log \log n)^{-\gamma_{f_X}/(2\beta_s)} e^{\alpha_{f_X}(\log \log n)^{\beta_{f_X}/\beta_s}/2},$$
$$\bar{\beta}_2(n) := (\log \log n)^{\max\{(\gamma_s - 2\beta_s + 3)/\beta_s, 0\}}.$$

(iv) $\beta_g = 0$ and $\beta_s = 0$. Suppose $\alpha_r, \beta_r > 0$. Then let $\bar{\zeta}_n = O((\log \log n)^{1/\beta_r})$ and $\bar{x}_n = O((\log \log \log n)^{1/\beta_{f_X}})$ to get

$$\sup_{y_2 \in \mathbb{Y}_2} \sup_{x_2^* \in \bar{\mathbb{X}}} \left| \check{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right| = O_p \left(\bar{\beta}(n)(\log n)^{-\alpha_r} \right) = o_p(1)$$

with

$$\bar{\beta}(n) := (\log \log n)^{(\gamma_r+3)/\beta_r} + (\log \log n)^{(2+\gamma_r-\gamma_g)/\beta_r}(\log \log \log n)^{(\gamma_s+3)/\beta_{f_X}}.$$

If the problems in (20) are severely ill-posed then the inversion of $\mathcal{D}_X$ is inherently difficult and leads to slow, logarithmic convergence rates when estimating the solutions to the corresponding ill-posed inverse problem. The logarithmic rates given in Corollary 4 reflect this well-known fact. Notice, however, that such slow rates occur also when the Fourier transform γ of the conditional cdf $F_{Y_2|X_2^*}$ decays rapidly relative to the characteristic function of the ME (Corollary 3(i),(ii)). To prevent the denominator from blowing up the estimation error, the estimator requires a lot of trimming in the sense that $\bar{\zeta}_n$ has to increase slowly. This trimming cuts off large parts of the integrand's tails, creates large biases, and leads to a slow overall convergence rate. In the remaining scenarios, the convergence rate is of polynomial order.

The next theorem establishes the convergence rates of the regression function estimators $\hat{m}$ as in Definition 8. Because the discussion in this section is restricted to univariate

mismeasured regressors, the following convergence rate only covers the model without fixed effects. However, the extension to accommodate fixed effects is straightforward as it only requires the analogous result for two dimensions, so that estimation can be based on differences.

Theorem 4. *Let $\bar{\mathbb{X}}$ be some compact subset of $\mathbb{R}$ and $\tilde{Y}_t = m(\tilde{X}_t^*) + \varepsilon_t$ for $t = 1, 2$. Suppose either $E[\varepsilon_t | \tilde{X}_t^*] = 0$, defining $\hat{m}$ as in (18), or $Q_{\varepsilon_t | \tilde{X}_t^*}(\tau | \tilde{x}_t^*) = 0$ for some $\tau \in (0, 1)$, defining $\hat{m}$ as in (19). Assume the distribution of $Y_2 \mid \tilde{X}_t^* = \tilde{x}_t^*$ has compact support for all values $\tilde{x}_t^*$. If any of the assumptions in the different cases of Corollaries 3 or 4 are satisfied then*

$$\sup_{\tilde{x}_2^* \in \bar{\mathbb{X}}} |\hat{m}(\tilde{x}_2^*) - m(\tilde{x}_2^*)| = O_p(\beta_m(n)),$$

where $\beta_m(n)$ denotes the convergence rate of $\check{F}_{Y_2 | X_2^*}$ defined in the relevant subcase of Corollary 3 or 4.

4 Simulations

This section studies the finite sample performance of the proposed estimator $\hat{m}$. I consider a nonlinear panel data regression without individual-specific heterogeneity,

$$Y_t = m(X_t^*) + \varepsilon_t, \qquad t = 1, 2,$$
$$X_t = X_t^* + \eta_t,$$

with the regression function $m(x^*) := \Phi(x^*) - 1/2$ and $\Phi(\cdot)$ the standard normal cdf. All variables are scalar. The latent true explanatory variables are generated by $X_1^* \sim N(0, 1)$ and $X_2^* = 0.8\,X_1^* + 0.7\,N(0, 1)$. The ME η_t is i.i.d. $N(0, \sigma_\eta^2)$ with $\sigma_\eta \in \{0.5, 1.5\}$, which correspond to the two scenarios called weak ME ($\sigma_\eta = 0.5$) and strong ME ($\sigma_\eta = 1.5$). The structural error is independent over time and drawn from $N(0, 1)$. The simulation results are based on $1,000$ Monte Carlo samples of length $n = 200$.

For the new ME-robust estimator, the simulation setup presents a worst-case scenario in the following sense. The choice of distributions implies (X_1, X_2) are jointly normal, resulting in severely ill-posed inverse problems. In addition, the conditional cdf $F_{Y_t | X_t^*}(y_t | \cdot)$ is super-smooth because the regression function and the density of the regression error have infinitely many derivatives. As shown in the previous section, this scenario leads to slow, logarithmic convergence rates.

The ill-posed inverse problem is regularized with $\alpha_n \in \{0.001, 0.01, 0.1\}$. The series estimator is based on a quadratic polynomial basis in x_2. For x_1, I consider polynomial

31

bases of orders $\{3, 5, 7\}$, cubic spline and cubic B-spline bases with $\{5, 10, 15\}$ knots. The function m is estimated on a grid of 128 equidistant x-values on $[-2, 2]$, which means that all discrete Fourier transforms involved are calculated at 128 values in the corresponding range in the frequency domain. The estimator $\hat{m}$ is computed as the conditional mean of $\check{F}_{Y_2|X_2^*}$, approximating the integral in (18) over a grid of 100 equidistant y-values between the 5% and the 95% quantile of all Y-draws pooled together. Other combinations of the various simulation parameters have been considered but did not have any qualitative effect on the results presented below.

Tables 1 and 2 summarize the performance of the new ME-robust estimator suggested in the previous section, compared with that of the standard Nadaraya-Watson estimator, which ignores the ME. The table reports the absolute value of bias, standard deviation, and root mean squared error, each averaged over the Monte Carlo samples and over the grid of 128 x-values. Figure 1 shows the estimated regression functions together with the range spanning two pointwise empirical standard deviations of the estimators.[13]

As mentioned in the previous section, the series estimator of the solution to the ill-posed inverse problems takes the form of a standard instrumental variable estimator for which the vectors of basis functions in x_1 and x_2 play the roles of the instruments and of the endogenous variables, respectively. To assess the reliability of this estimator, Table 3 reports the Cragg-Donald statistic of testing for weak instruments, which is a multivariate extension of the test based on the first-stage F-statistic; see Cragg and Donald (1993) and Stock and Yogo (2005) for details of the procedure.[14] The test is valid when the model is correctly specified, which, in the present context, requires the population model be expressed in terms of the finite basis vectors in **B** and **P**. In general, the finite-dimensional case can, of course, only approximate the truth. The table also lists the critical values for 5%-level tests of 10% two-stage least-squares bias and 15% two-stage least-squares size distortion. The Cragg-Donald test rejects weak identification for large values of the test statistic. Rejection occurs for the polynomial basis, whereas the results based on the (B-)spline basis appear less reliable.

In terms of bias, standard deviation, and root mean squared error, the results demonstrate similar performance of the new ME-robust and the standard Nadaraya-Watson estimator when ME is weak. In particular, the choices of tuning parameters (α_n, K_n, J_n) have almost no impact on the resulting estimates. However, when the ME has a large

---

[13]Notice the asymptotic distribution of the ME-robust estimator is not available, so the displayed measure of variability has to be interpreted with care.

[14]In practice, one would want to test completeness or some notion of weak completeness directly, but such a test may exist only in very special situations; see Canay, Santos, and Shaikh (2013).

variance, the Nadaraya-Watson estimator is strongly biased, whereas the bias of the ME-robust estimator barely changes relative to the weak ME scenario. This result is expected and confirms the theoretical finding that in the presence of ME, the ME-robust estimator is consistent, whereas the Nadaraya-Watson estimator is not. The variability of the ME-robust estimator is higher than (in the case of strong ME) or comparable to (in the case of weak ME) that of the Nadaraya-Watson estimator. This finding is not surprising either: just as in the case of linear regression with an endogenous regressor, the OLS estimator tends to have a smaller asymptotic variance than two-stage least squares, but the former is centered around the wrong quantity, whereas the latter is not. Similarly, here the Nadaraya-Watson estimator is biased and less variable, but only the ME-robust estimator consistently estimates the correct object. This simulation experiment confirms the theoretical findings from the previous sections in that the new estimator is indeed robust to ME and significantly reduces bias.

5   Conclusions

The paper presents a constructive identification argument for nonlinear panel data regressions with measurement error and fixed effects. The identifying assumptions are easy to interpret in the panel data context and resemble the standard conditions for identification of linear instrumental variable models. They inform the applied researcher about which types of panel data models are identified in the presence of measurement error, and what type of variation in the data is required to secure identification. I show that, under regularity conditions, if only outcomes and mismeasured regressors are observed, the model is identified if either the measurement error or the regression error is serially independent.

A   Sufficient Condition for Assumption IVX 1

Lemma 3. *Suppose Assumption ID 2(ii) holds, that $X_2^* = g(X_1^*) + U$ with $U \perp X_1^*$, that the characteristic function of U is nonzero on a dense subset of $\mathbb{R}^p$ and that $F_{Y_2|X_1}(y_2|\cdot)$ satisfies Assumption ID 3 with the obvious modifications. Then Assumption IVX 1 holds.*

Proof   The proof is straightforward. (i) $\mathcal{F}\mathcal{C}^{-1}\mathcal{F}^{-1}$ is equal to multiplication by the inverse of the characteristic function of the ME (see also proof of Theorem 1) which is nonzero. (ii) Similarly, $\mathcal{F}\mathcal{C}_{rev}^{-1}\mathcal{F}^{-1}$ is equivalent to multiplication by a nonzero function. (iii) By the assumption of the lemma, $\mathcal{T}$ is also a convolution operator whose kernel has a finite Fourier transform (the characteristic function of U). Therefore, $\mathcal{F}\mathcal{T}^{-1}\mathcal{F}^{-1}$ is also equal to a multiplication operator whose multiplicator is nonzero over whole $\mathbb{R}^p$. (iv) By Lemma 4(ii), $\mathcal{F}F_{Y_2|X_1}$ is nonzero as well.

In conclusion, (i) – (iv) above imply

$$\mathcal{F}\mathcal{D}_X^{-1}F_{Y_2|X_1} = \mathcal{F}\mathcal{C}^{-1}\mathcal{F}^{-1}\mathcal{F}\mathcal{T}^{-1}\mathcal{F}^{-1}\mathcal{F}\mathcal{C}_{rev}^{-1}\mathcal{F}^{-1}\mathcal{F}F_{Y_2|X_1}$$

is nonzero on a dense subset of $\mathbb{R}^p$.   Q.E.D.

B   Proofs

Constants in this section are denoted by C, C', C'', and so on, but the same symbol in different instances does not necessarily refer to the same value. Also, let λ denote Lebesgue measure on $\mathbb{R}^p$ and denote by $L^1(\lambda)$ the space of functions that are absolutely integrable. For two sequences $\{a_n\}$ and $\{b_n\}$ in $\mathbb{R}$, $a_n \asymp b_n$ means that the sequence a_n/b_n is bounded away from 0 and ∞ uniformly over n

B.1   Identification

Lemma 4. *Under Assumption ID 2(i), 3 and Assumption IVX 2(i), the following holds:*

(i) $\lim_{x_t \to \pm\infty} F_{Y_t|X_t}(y_t|x_t) = \lim_{x_t^* \to \pm\infty} F_{Y_t|X_t^*}(y_t|x_t^*)$.

(ii) *The function g as defined in the proof of Theorem 1 is not in $L^1(\lambda)$, but its Fourier transform γ is an ordinary function: $\gamma(\zeta) = [\mathcal{F}g_o](\zeta) - 2c_Y(y_2)/(\mathbf{i}\zeta)$ for $\zeta \in \mathbb{R}^p \setminus \{0\}$ and where g_o is a function in $L^1(\lambda)$.*

(iii) *The function $\tilde{g}$ as defined in the proof of Theorem 1 is not in $L^1(\lambda)$, but its Fourier transform is an ordinary function: $[\mathcal{F}\tilde{g}](\zeta) = [\mathcal{F}\tilde{g}_o](\zeta) - (c_Y^+(y_2) - c_Y^-(y_2))/|\zeta|^2$ for $\zeta \in \mathbb{R}^p \setminus \{0\}$ and where $\tilde{g}_o$ is a function in $L^1(\lambda)$.*

Proof Parts (i) and (ii) follow directly from Lemma 1 in Schennach (2008). A modification of that derivation proves part (iii) as follows. For simplicity of exposition, consider the univariate case $p = 1$. Define the function

$$H(x_2^*) := c_Y^-(y_2)\mathbb{1}\{x_2^* \leq 0\} + c_Y^+(y_2)\mathbb{1}\{x_2^* > 0\}$$

and note

$$\int_{-\infty}^0 \left| x_2^* \left(F_{Y_2|X_2^*}(y_2|x_2^*) - H(x_2^*) \right) \right| dx_2^*$$

$$= \int_{-\infty}^0 |x_2^*| \left| c_Y^-(y_2) + \int_{-\infty}^{x_2^*} \frac{\partial F_{Y_2|X_2^*}(y_2|u)}{\partial x_2^*} du - c_Y^-(y_2) \right| dx_2^*$$

$$\leq \int_{-\infty}^0 |x_2^*| \int_{-\infty}^{x_2^*} \left| \frac{\partial F_{Y_2|X_2^*}(y_2|u)}{\partial x_2^*} \right| du dx_2^*$$

$$\leq \int_{-\infty}^0 |x_2^*| \int_{-\infty}^{x_2^*} B(1 + |u|)^{-3-\beta} du dx_2^*$$

$$\leq B' \int_{-\infty}^0 |x_2^*|(1 + |x_2^*|)^{-2-\beta} dx_2^*$$

$$\leq B' \int_{-\infty}^0 |x_2^*|(1 + |x_2^*|)^{-2-\beta} dx_2^* + B' \int_{-\infty}^0 (1 + |x_2^*|)^{-2-\beta} dx_2^*$$

$$- B' \int_{-\infty}^0 (1 + |x_2^*|)^{-2-\beta} dx_2^*$$

$$\leq B' \int_{-\infty}^0 (1 + |x_2^*|)^{-1-\beta} dx_2^* - B' \int_{-\infty}^0 (1 + |x_2^*|)^{-2-\beta} dx_2^* < \infty.$$

Similarly, one can show the same integral from 0 to ∞ is finite. Then, defining $\tilde{g}_o(x_2^*) := x_2^* \left(F_{Y_2|X_2^*}(y_2|x_2^*) - H(x_2^*) \right)$, we have that $\int |\tilde{g}_o(x_2^*)| dx_2^* < \infty$. Furthermore, $\tilde{g}(x_2^*) := x_2^*(F_{Y_2|X_2^*}(y_2|x_2^*) - c_Y(y_2))$ can be decomposed as

$$\tilde{g}(x_2^*) = x_2^* \left(F_{Y_2|X_2^*}(y_2|x_2^*) - H(x_2^*) \right) + x_2^* \left(H(x_2^*) - c_Y(y_2) \right)$$

$$= \tilde{g}_o(x_2^*) + x_2^* \left(c_Y^-(y_2)\mathbb{1}\{x_2^* \leq 0\} + c_Y^+(y_2)\mathbb{1}\{x_2^* > 0\} - c_Y(y_2) \right)$$

$$= \tilde{g}_o(x_2^*) + x_2^* \left(\frac{c_Y^-(y_2)}{2} [2\mathbb{1}\{x_2^* \leq 0\} - 1] + \frac{c_Y^+(y_2)}{2} [2\mathbb{1}\{x_2^* > 0\} - 1] \right)$$

$$= \tilde{g}_o(x_2^*) + x_2^* \left(-\frac{c_Y^-(y_2)}{2} [2\mathbb{1}\{x_2^* > 0\} - 1] + \frac{c_Y^+(y_2)}{2} [2\mathbb{1}\{x_2^* > 0\} - 1] \right)$$

$$= \tilde{g}_o(x_2^*) + x_2^* \frac{c_Y^+(y_2) - c_Y^-(y_2)}{2} sgn(x_2^*)$$

$$= \tilde{g}_o(x_2^*) + (c_Y(y_2) - c_Y^-(y_2)) |x_2^*|.$$

where the first term is absolutely integrable as shown above, and $(c_Y(y_2) - c_Y^-(y_2))|x_2^*|$ is not but possesses the Fourier transform $\zeta \mapsto -2(c_Y(y_2) - c_Y^-(y_2))/|\zeta|^2$. The case $p > 1$ follows from a similar argument. Q.E.D.

Lemma 5. *Under Assumptions ID 1 and 2, $\mathcal{C} : L^2(f_{X_2}) \to L^2(f_{X_2^*})$ and $\mathcal{C}_{rev} : L^2(f_{X_1^*}) \to L^2(f_{X_1})$ are injective. Furthermore, $\mathcal{D}_X$ is injective on $\mathcal{C}^{-1}(\mathcal{G})$ and $\mathcal{D}_X \mathcal{C}^{-1} h = \mathcal{C}_{rev} \mathcal{T} h$ holds for all $h \in \mathcal{G}$.*

Proof The injectivity of $\mathcal{C}$ and $\mathcal{C}_{rev}$ is straightforward to prove. See, for example, Proposition 8 of Carrasco and Florens (2011). By similar reasoning, the adjoint operator $\mathcal{C}^*$ of $\mathcal{C}$ is injective as well. Since $\mathcal{C}$ is injective, it possesses an inverse over its range. Since its adjoint $\mathcal{C}^*$ is injective as well and $\overline{R(\mathcal{C})} = N(\mathcal{C}^*)^\perp = L^2(f_{X_2^*})$, where $R(\cdot)$ and $N(\cdot)$ denote the range and null space of an operator, the range of $\mathcal{C}$ is dense in $L^2(f_{X_2^*})$ and $\mathcal{D}_X \mathcal{C}^{-1} = \mathcal{C}_{rev} \mathcal{T}$ holds over whole $L^2(f_{X_2^*})$. Therefore, $\mathcal{D}_X \mathcal{C}^{-1} = \mathcal{C}_{rev} \mathcal{T}$ holds over $\mathcal{G}$ in particular. By Assumption 4, $\mathcal{T}$ is invertible over $\mathcal{G}$ and thus $\mathcal{D}_X$ is invertible over $\mathcal{C}^{-1}(\mathcal{G})$, which concludes the proof of the lemma. Q.E.D.

Proof of Theorem 1 First, notice that by Assumption IVX 2(i) and (ii),

$$d_Y(x_1) := E[\mathbb{1}\{Y_2 \le y_2\}|X_1 = x_1]$$

$$= \iint F_{Y_2|X_2^*,X_1^*,X_1}(y_2|x_2^*, x_1^*, x_1) f_{X_2^*,X_1^*|X_1}(x_2^*, x_1^*|x_1) dx_2^* dx_1^*$$

$$= \iint F_{Y_2|X_2^*}(y_2|x_2^*) f_{X_2^*|X_1^*}(x_2^*|x_1^*) f_{X_1^*|X_1}(x_1^*|x_1) dx_2^* dx_1^* \qquad (21)$$

and

$$d_{YX}(x_1) := E[X_2 \mathbb{1}\{Y_2 \le y_2\}|X_1 = x_1]$$

$$= \int E[X_2 \mathbb{1}\{Y_2 \le y_2\}|X_2^* = x_2^*] f_{X_2^*|X_1}(x_2^*|x_1) dx_2^*$$

$$= \int E[X_2|X_2^* = x_2^*] F_{Y_2|X_2^*}(y_2|x_2^*) f_{X_2^*|X_1}(x_2^*|x_1) dx_2^*$$

$$= \iint x_2^* F_{Y_2|X_2^*}(y_2|x_2^*) f_{X_2^*|X_1^*}(x_2^*|x_1^*) f_{X_1^*|X_1}(x_1^*|x_1) dx_2^* dx_1^*. \qquad (22)$$

Letting $\tilde{F}_{Y_2|X_2^*}(y_2|x_2^*) := x_2^* F_{Y_2|X_2^*}(y_2|x_2^*)$ and with the operators introduced in the main text, these two equations can be rewritten as

$$d_Y = \mathcal{C}_{rev}\mathcal{T}F_{Y_2|X_2^*}, \tag{23}$$

$$d_{YX} = \mathcal{C}_{rev}\mathcal{T}\tilde{F}_{Y_2|X_2^*}. \tag{24}$$

Here, as well as in the remainder of the identification argument, I suppress the dependence of various functions on y_2 as this value is fixed throughout. Next, notice that by Assumption IVX 2, we have

$$f(x_2|x_1) = \iint f_{X_2|X_2^*}(x_2|x_2^*) f_{X_2^*|X_1^*}(x_2^*|x_1^*) f_{X_1^*|X_1}(x_1^*|x_1) dx_2^* dx_1^*$$

which can also be written as the operator identity $\mathcal{D}_X = \mathcal{C}_{rev}\mathcal{T}\mathcal{C}$, i.e. $\mathcal{D}_X h = \mathcal{C}_{rev}\mathcal{T}\mathcal{C}h$ for $h \in L^2(f_{X_2})$. By Lemma 5, $\mathcal{C}$ is invertible so that substituting the expression for $\mathcal{D}_X$ into (23) and (24) yields

$$d_Y = \mathcal{D}_X\mathcal{C}^{-1}F_{Y_2|X_2^*}, \tag{25}$$

$$d_{YX} = \mathcal{D}_X\mathcal{C}^{-1}\tilde{F}_{Y_2|X_2^*}. \tag{26}$$

By Lemma 5, $\mathcal{D}_X\mathcal{C}^{-1}h = \mathcal{C}_{rev}\mathcal{T}h$ for any $h \in \mathcal{G}$, a space of functions on which $\mathcal{T}$ and thus $\mathcal{D}_X$ are injective. Therefore, we have

$$\mathcal{C}\mathcal{D}_X^{-1}d_Y = F_{Y_2|X_2^*}, \tag{27}$$

$$\mathcal{C}\mathcal{D}_X^{-1}d_{YX} = \tilde{F}_{Y_2|X_2^*}. \tag{28}$$

Next, I show a unique solution $F_{Y_2|X_2^*}$ to these two equations exists. Let $s_Y(x_2) := [\mathcal{D}_X^{-1}d_Y](x_2) - c_Y(y_2)$ and $s_{YX}(x_2) := [\mathcal{D}_X^{-1}d_{YX}](x_2) - x_2 c_Y(y_2)$ and define the centered counterparts of $F_{Y_2|X_2^*}(y_2|x_2^*)$ and $\tilde{F}_{Y_2|X_2^*}(y_2|x_2^*)$ as $g(x_2^*) := F_{Y_2|X_2^*}(y_2|x_2^*) - c_Y(y_2)$ and $\tilde{g}(x_2^*) := \tilde{F}_{Y_2|X_2^*}(y_2|x_2^*) - x_2^* c_Y(y_2)$. The function c_Y centers the various functions involved such that by Lemma 4, g and $\tilde{g}$ possess Fourier transforms that are ordinary functions. Formally, if the covariates are scalars $(p = 1)$, define $c_Y(y_2) := (c_Y^+(y_2) + c_Y^-(y_2))/2$ with

$$c_Y^+(y_2) := \lim_{x_2 \to \infty} F_{Y_2|X_2}(y_2|x_2),$$

$$c_Y^-(y_2) := \lim_{x_2 \to -\infty} F_{Y_2|X_2}(y_2|x_2).$$

If $p > 1$, c_Y can be selected as

$$c_Y(y_2) := \lim_{R_1 \to \infty} \lim_{R_2 \to \infty} \frac{\int_{R_1 \le |x| \le R_2} F_{Y_t|X_t}(y_t|x)dx}{\int_{R_1 \le |x| \le R_2} dx}.$$

Subtracting the centering constants and using $E[X_2|X_2^* = x_2^*] = x_2^*$, (27) and (28) are equivalent to

$$\mathcal{C}s_Y = g, \tag{29}$$

$$\mathcal{C}s_{YX} = \tilde{g}. \tag{30}$$

Next, let $\mathcal{F}$ denote the Fourier transform operator $[\mathcal{F}h](\zeta) := \int h(x)e^{i\zeta \cdot x}dx$, $\zeta \in \mathbb{R}^p$. It is well-known that $\mathcal{F}$ diagonalizes convolution operators[15] such as $\mathcal{C}$; that is, $\mathcal{F}\mathcal{C}\mathcal{F}^{-1} = \Delta_\phi$ with the multiplication operator $[\Delta_\phi h](\zeta) := \phi(\zeta)h(\zeta)$ and ϕ the characteristic function of $-\eta_2$. Therefore, $\mathcal{F}\mathcal{C}s_Y = \mathcal{F}\mathcal{C}\mathcal{F}^{-1}\mathcal{F}s_Y = \phi\,\sigma_Y$ and, similarly, $\mathcal{F}\mathcal{C}s_{YX} = \phi\,\sigma_{YX}$, where σ_Y and σ_{YX} are the Fourier transforms of s_Y and s_{YX}. Therefore, applying $\mathcal{F}$ to both (29) and (30) yields

$$\phi(\zeta)\,\sigma_Y(\zeta) = \gamma(\zeta), \qquad \zeta \neq 0, \tag{31}$$

$$\phi(\zeta)\,\sigma_{YX}(\zeta) = -i\nabla\gamma(\zeta), \qquad \zeta \neq 0. \tag{32}$$

The last equality holds because multiplication by ix_2^* corresponds to taking derivatives in the Fourier domain. The equations are valid for all $\zeta \neq 0$ because, by Lemma 4, the Fourier transform γ and its partial derivatives have poles at the origin but are ordinary functions everywhere else. Now, differentiate (31) with respect to ζ, substitute in (32), and divide by ϕ (which is allowed by Assumption ID 2(ii)) to get

$$\frac{\nabla_\zeta\phi(\zeta)}{\phi(\zeta)}\sigma_Y(\zeta) + \nabla_\zeta\sigma_Y(\zeta) = \frac{\nabla_\zeta\gamma(\zeta)}{\phi(\zeta)} = i\sigma_{YX}(\zeta), \qquad \zeta \neq 0$$

or the following set of partial differential equations in ϕ:

$$\frac{\nabla_\zeta\phi(\zeta)}{\phi(\zeta)} = \frac{i\sigma_{YX}(\zeta) - \nabla_\zeta\sigma_Y(\zeta)}{\sigma_Y(\zeta)}, \qquad \zeta \neq 0.$$

This equation holds for all $\zeta \in \{\zeta \neq 0 : \sigma_Y(\zeta) \neq 0\}$, but since the left-hand side is an ordinary continuous function on whole $\mathbb{R}^p$ and by Assumption IVX 1, one can uniquely extend the quotient on the right to $\mathbb{R}^p$ by a continuous limiting process. Subsequently, let $\sigma := (i\sigma_{YX} - \nabla_\zeta\sigma_Y)/\sigma_Y$ denote this extension. Solving the partial differential equations with the initial condition $\phi(0) = 1$ yields

$$\phi(\zeta) = \exp\left\{\int_0^\zeta \sigma(z) \cdot d\Gamma(z)\right\},$$

---

[15]See, for example, section 3 of Carroll, Rooij, and Ruymgaart (1991).

where the integral is a path integral along some smooth path Γ that connects 0 and ζ in $\mathbb{R}^p$. Then, by equation (31), g is identified as the inverse Fourier transform

$$g(x_2^*) = \frac{1}{2\pi}\int \phi(\zeta)\sigma_Y(\zeta)e^{-i\zeta\cdot x_2^*}d\zeta$$

and $F_{Y_2|X_2^*}(y_2|x_2^*) = g(x_2^*) + c_Y(y_2)$.

The marginal distribution of the ME η_2 is identified from the inverse Fourier transform of the characteristic function ϕ. Knowledge of the marginal ME distribution in the second period implies identification of the contamination operator $\mathcal{C}$. If the distribution of the ME is stationary, then $\mathcal{C}_{rev}$ can be calculated from $\mathcal{C}$, yielding an expression of the transition law $\mathcal{T}$ via $\mathcal{T} = \mathcal{C}_{rev}^{-1}\mathcal{D}_X\mathcal{C}^{-1}$. Therefore, the distribution of X_2^* given X_1^* is known. From the characteristic function $E[e^{i\zeta\cdot X_1^*}] = E[e^{i\zeta\cdot X_1}]/E[e^{i\zeta\cdot\eta_1}]$, we then get the marginal distribution of X_1^*, which, together with the transition law, identifies the joint distribution of X^* as well as that of η (from $E[e^{i\zeta\cdot\eta}] = E[e^{i\zeta\cdot X^*}]/E[e^{i\zeta\cdot X^*}]$).   Q.E.D.

Proof of Corollary 1   The corollary follows from the proof of Theorem 1.   Q.E.D.

Proof of Theorem 2   To simplify the exposition, suppose $p = 2$ and $T = 3$. The argument for larger dimensions p and $T = p + 1$ works analogously; if $T > p + 1$, identification of $F_{Y_T|X_T^*}$ can be based only on the last $p + 1$ periods. Consider the two equations

$$\begin{aligned}
d_Y^{\circ}(y_{1:2}) &:= E[\mathbb{1}\{Y_3 \leq y_3\}|Y_{1:2} = y_{1:2}]\\
&= \iint F_{Y_3|X_3^*,X_2^*,Y_{1:2}}(y_3|x_3^*,x_2^*,y_{1:2})f_{X_3^*|X_2^*,Y_{1:2}}(x_3^*|x_2^*,y_{1:2})f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_3^*dx_2^*\\
&= \iint F_{Y_3|X_3^*}(y_3|x_3^*)f_{X_3^*|X_2^*}(x_3^*|x_2^*)f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_3^*dx_2^*
\end{aligned}$$

(33)

and

$$\begin{aligned}
d_{YX}^{\circ}(y_{1:2}) &:= E[X_3\mathbb{1}\{Y_3 \leq y_3\}|Y_{1:2} = y_{1:2}]\\
&= \iint x_3^*F_{Y_3|X_3^*}(y_3|x_3^*)f_{X_3^*|X_2^*}(x_3^*|x_2^*)f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_3^*dx_2^*,
\end{aligned}$$

(34)

which hold by Assumption IVY 2. Letting $\tilde{F}_{Y_3|X_3^*}(y_3|x_3^*) := x_3^*F_{Y_3|X_3^*}(y_3|x_3^*)$ and with the operators introduced in the main text, these two equations are equivalent to

$$d_Y^{\circ} = \mathcal{M}^*\mathcal{T}F_{Y_3|X_3^*},$$

(35)

$$d_{YX}^{\circ} = \mathcal{M}^*\mathcal{T}\tilde{F}_{Y_3|X_3^*},$$

(36)

keeping $y_3 \in \mathbb{R}$ fixed and implicit for the remainder of the proof. Similarly,

$$
\begin{aligned}
f_{X_3|Y_{1:2}}&(x_3|y_{1:2}) \\
&= \iint f_{X_3|X_3^*,X_2^*,Y_{1:2}}(x_3|x_3^*,x_2^*,y_{1:2})f_{X_3^*|X_2^*,Y_{1:2}}(x_3^*|x_2^*,y_{1:2})f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_3^*dx_2^* \\
&= \iint f_{X_3|X_3^*}(x_3|x_3^*)f_{X_3^*|X_2^*}(x_3^*|x_2^*)f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_3^*dx_2^*,
\end{aligned}
\tag{37}
$$

which is equivalent to $\mathcal{D}_Y = \mathcal{M}^*\mathcal{T}\mathcal{C}$. Next, we need to show that $\mathcal{D}_Y$ is invertible on $\mathcal{C}^{-1}(\mathcal{G})$. As in the proof of Theorem 1, $\mathcal{C}$ is a convolution operator whose kernel has a nonzero Fourier transform and therefore is invertible. $\mathcal{T}$ is invertible by Assumption ID 4. Therefore, it remains to show $\mathcal{M}^*$ is invertible on its range. To that end, denote by $\mathcal{B}$ the set of all bounded functions from $\mathbb{R}^p$ to $\mathbb{R}^p$ and notice, by Assumption IVY 3(iii), $\mathcal{T}$ maps $\mathcal{G}$ into $\mathcal{B}$. Therefore, we only need to show bounded completeness of the conditional distribution of $X_2^*|(Y_1, Y_2)$. Let $h \in \mathcal{B}$ and consider

$$
\begin{aligned}
\int h(x_2^*)f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_2^* &= \iint h(x_2^*)f_{X_1^*,X_2^*|Y_{1:2}}(x_1^*,x_2^*|y_{1:2})dx_1^*dx_2^* \\
&= \iint h(x_2^*)\frac{f_{X_1^*,X_2^*}(x_1^*,x_2^*)}{f_{Y_{1:2}}(y_{1:2})}f_{Y_1|X_1^*}(y_1|x_1^*)f_{Y_2|X_2^*}(y_2|x_2^*)dx_1^*dx_2^* \\
&= \iint h(x_2^*)\frac{f_{X_1^*,X_2^*}(x_1^*,x_2^*)}{f_{Y_{1:2}}(y_{1:2})}f_{\varepsilon_1}(y_1 - g(x_1^*))f_{\varepsilon_2}(y_2 - g(x_2^*))dx_1^*dx_2^*.
\end{aligned}
$$

Define $\tilde{h}(u_1, u_2) := h(m^{-1}(u_2))f_{X_1^*,X_2^*}(m^{-1}(u_1), m^{-1}(u_2))$. By the previous equation and Assumption IVY 3(ii), setting $\int h(x_2^*)f_{X_2^*|Y_{1:2}}(x_2^*|y_{1:2})dx_2^*$ to zero is equivalent to setting

$$
\begin{aligned}
\iint \tilde{h}(u_1, u_2)&\frac{f_{\varepsilon_1}(y_1 - u_1)f_{\varepsilon_2}(y_2 - u_2)}{f_{Y_{1:2}}(y_{1:2})}du_1du_2 \\
&= \frac{1}{f_{Y_{1:2}}(y_{1:2})}\int\left(\int \tilde{h}(u_1, u_2)f_{\varepsilon_1}(y_1 - u_1)du_1\right)f_{\varepsilon_2}(y_2 - u_2)du_2
\end{aligned}
\tag{38}
$$

to zero. By Assumption IVY 3(ii), $f_{Y_{1:2}}(y_{1:2}) \neq 0$ for all $y_{1:2} \in \mathbb{R}^2$. By Assumptions IVY 3(i), (ii), and (iv), the integrals with respect to $f_{\varepsilon_1}(y_1 - u_1)$ and $f_{\varepsilon_2}(y_2 - u_2)$ are convolutions with nonzero Fourier transform of their respective kernels. Furthermore, $\int \tilde{h}(u_1, u_2)f_{\varepsilon_1}(y_1 - u_1)du_1$ is a bounded function in u_2, and so (38) implies that $\int \tilde{h}(u_1, m(X_2^*))f_{\varepsilon_1}(y_1 - u_1)du_1 = 0$ a.s. whenever $E[h(X_2^*)|Y_{1:2}] = 0$ a.s.. Similarly, for any $u_2 \in \mathbb{R}^p$, $\tilde{h}(\cdot, u_2) = 0$ whenever $\int \tilde{h}(u_1, u_2)f_{\varepsilon_1}(\cdot - u_1)du_1 = 0$. By Assumption IVY 3(ii), f_{X^*} is positive everywhere so, in conclusion, we have $E[h(X_2^*)|Y_{1:2}] = 0$ a.s. implies $h(X_2^*) = 0$ a.s., the desired completeness result.

Having established invertibility of $\mathcal{D}_Y$, use the relationship $\mathcal{D}_Y = \mathcal{M}^* \mathcal{T} \mathcal{C}$ to rewrite (35) and (36) as

$$\mathcal{C} \mathcal{D}_Y^{-1} d_Y^\circ = F_{Y_3|X_3^*},$$
$$\mathcal{C} \mathcal{D}_Y^{-1} d_{YX}^\circ = \tilde{F}_{Y_3|X_3^*},$$

and the remainder of the proof closely follows that of Theorem 1. Q.E.D.

Proof of Corollary 2 The corollary follows from the proof of Theorem 2. Q.E.D.

Proof of Lemma 1 To simplify the exposition, I subsequently drop the arguments of conditional densities that should be obvious from the context.

First, by Assumption REG 1(i) and (ii), $(\varepsilon_3, \varepsilon_4) \perp (\tilde{\eta}_1, \tilde{\eta}_2) \mid \tilde{X}^*$ and $(\varepsilon_3, \varepsilon_4) \perp (\tilde{X}_1^*, \tilde{X}_2^*) \mid (\tilde{X}_3^*, \tilde{X}_4^*)$. These two independence conditions imply $f_{\varepsilon_3, \varepsilon_4 | \tilde{X}^*, \tilde{\eta}_1, \tilde{\eta}_2} = f_{\varepsilon_3, \varepsilon_4 | \tilde{X}^*} = f_{\varepsilon_3, \varepsilon_4 | \tilde{X}_3^*, \tilde{X}_4^*}$ and thus $f_{\Delta \varepsilon_4 | \tilde{X}^*, \tilde{\eta}_1, \tilde{\eta}_2} = f_{\Delta \varepsilon_4 | \tilde{X}_3^*, \tilde{X}_4^*}$. Therefore,

$$f_{Y_2|X_2^*, X_1^*, X_1} = f_{\Delta \tilde{Y}_4 | \tilde{X}^*, \tilde{X}_2, \tilde{X}_1} = f_{\Delta \varepsilon_4 | \tilde{X}^*, \tilde{X}_2, \tilde{X}_1} = f_{\Delta \varepsilon_4 | \tilde{X}^*, \tilde{\eta}_2, \tilde{\eta}_1}$$
$$= f_{\Delta \varepsilon_4 | \tilde{X}_3^*, \tilde{X}_4^*} = f_{\Delta \tilde{Y}_4 | \tilde{X}_3^*, \tilde{X}_4^*} = f_{Y_2|X_2^*}. \tag{39}$$

Second, Assumption ID 2(i) and Assumption REG 1(iii) imply the weaker statements $(\tilde{\eta}_3, \tilde{\eta}_4) \perp (\tilde{\eta}_1, \tilde{\eta}_2) \mid \tilde{X}^*$ and $(\tilde{\eta}_3, \tilde{\eta}_4) \perp (\tilde{X}_1^*, \tilde{X}_2^*) \mid (\tilde{X}_3^*, \tilde{X}_4^*)$ so that $f_{\tilde{\eta}_3, \tilde{\eta}_4 | \tilde{X}^*, \tilde{\eta}_1, \tilde{\eta}_2} = f_{\tilde{\eta}_3, \tilde{\eta}_4 | \tilde{X}^*} = f_{\tilde{\eta}_3, \tilde{\eta}_4 | \tilde{X}_3^*, \tilde{X}_4^*}$. Therefore,

$$f_{X_2|X_2^*, X_1^*, X_1} = f_{\tilde{X}_3, \tilde{X}_4 | \tilde{X}^*, \tilde{X}_2, \tilde{X}_1} = f_{\tilde{\eta}_3, \tilde{\eta}_4 | \tilde{X}^*, \tilde{X}_2, \tilde{X}_1} = f_{\tilde{\eta}_3, \tilde{\eta}_4 | \tilde{X}^*, \tilde{\eta}_2, \tilde{\eta}_1}$$
$$= f_{\tilde{\eta}_3, \tilde{\eta}_4 | \tilde{X}_3^*, \tilde{X}_4^*} = f_{\tilde{X}_3, \tilde{X}_4 | \tilde{X}_3^*, \tilde{X}_4^*} = f_{X_2|X_2^*}. \tag{40}$$

Third, by Assumption REG 1(ii),

$$f_{Y_2|X_2^*, X_1^*, X_2, X_1} = f_{\Delta \tilde{Y}_4 | \tilde{X}^*, \tilde{X}} = f_{\Delta \varepsilon_4 | \tilde{X}^*, \tilde{\eta}} = f_{\Delta \varepsilon_4 | \tilde{X}^*, \tilde{\eta}_1, \tilde{\eta}_2} = f_{\Delta Y_4 | \tilde{X}^*, \tilde{X}_1, \tilde{X}_2} = f_{Y_2|X_2^*, X_1^*, X_1} \tag{41}$$

Now, (41) implies Assumption IVX 2(ii), which in turn means (39) and (40) together imply Assumption IVX 2(i). Therefore, Theorem 1 can be applied to identify the conditional cdf $F_{\Delta \tilde{Y}_4 | \tilde{X}_4^*, \tilde{X}_3^*}$. Because of Assumption REG 2,

$$E[\Delta \tilde{Y}_4 | \tilde{X}_4^* = \tilde{x}_4^*, \tilde{X}_3^* = 0] = m(\tilde{x}_4^*) - m(0) + E[\varepsilon_4 | \tilde{X}_4^* = \tilde{x}_4^*, \tilde{X}_3^* = 0]$$
$$- E[\varepsilon_3 | \tilde{X}_4^* = \tilde{x}_4^*, \tilde{X}_3^* = 0]$$
$$= m(\tilde{x}_4^*) - m(0) - E[\varepsilon_3 | \tilde{X}_3^* = 0],$$

so the regression function m can be written as

$$m(\tilde{x}_4^*) = const. + \int \Delta y \, dF_{\Delta \tilde{Y}_4 | \tilde{X}_4^*, \tilde{X}_3^*}(\Delta y | \tilde{x}_4^*, 0)$$

and the statement of the lemma follows. Q.E.D.

Proof of Lemma 2 Analogous to the proof of Lemma 1. Q.E.D.

B.2   Consistency and Convergence Rates

Proof of Theorem 3 The derivation of the convergence rate proceeds in roughly four steps: (i) bound $\|\hat{s}_Y - s_Y\|$ and similar estimation errors of the other s-functions; (ii) use step (i) to bound $\|\hat{\sigma}_Y - \sigma_Y\|$ and similar estimation errors for the other σ-functions; (iii) use step (ii) to bound $\|\hat{\sigma}_\Delta / \hat{\sigma}_Y - \sigma_\Delta / \sigma_Y\|$, where σ_Δ ($\hat{\sigma}_\Delta$) is (an estimator of) the difference in two of the σ-functions; (iv) use the previous steps to get the desired bound on the estimation error in $\check{F}_{Y_t | X_t^*}$.

Step (i) By Theorem 2 of Blundell, Chen, and Kristensen (2007), we have $\|\hat{s}_Y - s_Y\| = O_p(\delta_n)$ and $\|\hat{s}_{YX} - s_{YX}\| = O_p(\delta_n)$ with $\delta_n := K_n^{\rho_s} + \tau_n \sqrt{K_n / n}$. These rates hold pointwise for a fixed y_2 that is kept implicit in the notation. Similarly, the derivatives can be estimated at the rates $\|\nabla^d \hat{s}_Y - \nabla^d s_Y\| = O_p(\omega_{d,n} \delta_n)$ and $\|\nabla^d \hat{s}_{YX} - \nabla^d s_{YX}\| = O_p(\omega_{d,n} \delta_n)$, for $d = 1, 2$, which follows from going through Blundell, Chen, and Kristensen (2007)'s proof and applying Newey (1997)'s Theorem 1 with $d = 1, 2$ instead of $d = 0$.

Step (ii) Consider the estimation error of the Fourier transform $\hat{\sigma}_Y$ and decompose it as follows:

$$\sup_{|\zeta| \leq \bar{\zeta}_n} |\mathbf{i}\zeta \hat{\sigma}_Y(\zeta) - \mathbf{i}\zeta \sigma_Y(\zeta)| \leq \sup_{|\zeta| \leq \bar{\zeta}_n} |\mathbf{i}\zeta \hat{\sigma}_Y(\zeta) - \mathbf{i}\zeta [\mathcal{F}\hat{s}_Y](\zeta)| + \sup_{|\zeta| \leq \bar{\zeta}_n} |\mathbf{i}\zeta [\mathcal{F}\hat{s}_Y](\zeta) - \mathbf{i}\zeta \sigma_Y(\zeta)|.$$

Consider each of the two terms separately. First,

$$\sup_{|\zeta|\leq\bar\zeta_n}\left|\mathbf{i}\zeta\hat\sigma_Y(\zeta)-\mathbf{i}\zeta[\mathcal{F}\hat s_Y](\zeta)\right|=\sup_{|\zeta|\leq\bar\zeta_n}\left|\mathbf{i}\zeta\int\left[\check s_Y(x_2)-\hat s_Y(x_2)\right]e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$=\sup_{|\zeta|\leq\bar\zeta_n}\left|\mathbf{i}\zeta\int_{|x_2|>\bar x_n}\left[\check s_Y(x_2)-\hat s_Y(x_2)\right]e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\leq\sup_{|\zeta|\leq\bar\zeta_n}\left|\mathbf{i}\zeta\int_{|x_2|>\bar x_n}\left[\check s_Y(x_2)-s_Y(x_2)\right]e^{\mathbf{i}\zeta x_2}dx_2\right|+\sup_{|\zeta|\leq\bar\zeta_n}\left|\mathbf{i}\zeta\int_{|x_2|>\bar x_n}\left[\hat s_Y(x_2)-s_Y(x_2)\right]e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\leq\bar\zeta_n\int_{|x_2|>\bar x_n}\left|\check s_Y(x_2)-s_Y(x_2)\right|dx_2+\sup_{|\zeta|\leq\bar\zeta_n}\left|\int\left[\nabla\hat s_Y(x_2)-\nabla s_Y(x_2)\right]e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\leq\int\left|\nabla\hat s_Y(x_2)-\nabla s_Y(x_2)\right|dx_2+O(T_{Y,n}^{d=0,j=0,k=1})$$

$$\leq\int_{|x_2|\leq\bar x_n}\left|\nabla\hat s_Y(x_2)-\nabla s_Y(x_2)\right|dx_2+\int_{|x_2|>\bar x_n}\left|\nabla\hat s_Y(x_2)-\nabla s_Y(x_2)\right|dx_2+O(T_{Y,n}^{d=0,j=0,k=1})$$

$$\leq\left[\frac{1}{\inf_{|x_2|\leq\bar x_n}|f_{X_2}(x_2)|}\int_{|x_2|\leq\bar x_n}\left|\nabla\hat s_Y(x_2)-\nabla s_Y(x_2)\right|^2 f_{X_2}(x_2)dx_2\right]^{1/2}$$

$$+O(T_{Y,n}^{d=0,j=0,k=1}+T_{Y,n}^{d=1,j=0,k=0})$$

$$=\underline{f}_n^{-1/2}\|\nabla\hat s_Y-\nabla s_Y\|+O(T_{Y,n}^{d=0,j=0,k=1}+T_{Y,n}^{d=1,j=0,k=0})$$

$$=O_p(\underline{f}_n^{-1/2}\delta_n\omega_{1,n})+O(T_{Y,n}^{d=0,j=0,k=1}+T_{Y,n}^{d=1,j=0,k=0}).$$

Similarly, the second term is bounded by

$$\sup_{|\zeta|\leq\bar\zeta_n}\left|\mathbf{i}\zeta[\mathcal{F}\hat s_Y](\zeta)-\mathbf{i}\zeta\sigma_Y(\zeta)\right|=\sup_{|\zeta|\leq\bar\zeta_n}\left|\int\left[\nabla\hat s_Y(x_2)-\nabla s_Y(x_2)\right]e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$=O_p(\underline{f}_n^{-1/2}\delta_n\omega_{1,n})+O(T_{Y,n}^{d=0,j=0,k=1}+T_{Y,n}^{d=1,j=0,k=0})$$

so that, in conclusion, letting $T_{Y,n}:=T_{Y,n}^{0,0,1}+T_{Y,n}^{1,0,0}$,

$$\epsilon_n:=\sup_{|\zeta|\leq\bar\zeta_n}\left|\hat\sigma_Y(\zeta)-\sigma_Y(\zeta)\right|=O_p(\underline{f}_n^{-1/2}\delta_n\omega_{1,n}\bar\zeta_n^{-1})+O(T_{Y,n}). \tag{42}$$

The next estimation error to bound is that of the numerator in the expression of $\phi(\zeta)$. To that end, let $\sigma_\Delta(\zeta):=\mathbf{i}\sigma_{YX}(\zeta)-\nabla\sigma_Y(\zeta)$ and $\hat\sigma_\Delta(\zeta):=\mathbf{i}\hat\sigma_{YX}(\zeta)-\nabla\hat\sigma_Y(\zeta)$. Consider

$$\sup_{|\zeta|\leq\bar\zeta_n}\left|\zeta^2\hat\sigma_\Delta(\zeta)-\zeta^2\sigma_\Delta(\zeta)\right|=\sup_{|\zeta|\leq\bar\zeta_n}\left|\zeta^2\hat\sigma_\Delta(\zeta)-\zeta^2[\mathcal{F}\hat s_\Delta](\zeta)\right|+\sup_{|\zeta|\leq\bar\zeta_n}\left|\zeta^2[\mathcal{F}\hat s_\Delta](\zeta)-\zeta^2\sigma_\Delta(\zeta)\right|,$$

43

where $\hat{s}_\Delta(x_2) := \check{s}_{YX}(x_2) - \mathbf{i}x_2\check{s}_Y(x_2)$. The first term can be bounded as follows:

$$\sup_{|\zeta|\le\bar{\zeta}_n} \left|\zeta^2\hat{\sigma}_\Delta(\zeta) - \zeta^2[\mathcal{F}\hat{s}_\Delta](\zeta)\right|$$

$$= \sup_{|\zeta|\le\bar{\zeta}_n} \left|\zeta^2 \int \left[\mathbf{i}\check{s}_{YX}(x_2) - \mathbf{i}x_2\check{s}_Y(x_2) - (\mathbf{i}\hat{s}_{YX}(x_2) - \mathbf{i}x_2\hat{s}_Y(x_2))\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$= \sup_{|\zeta|\le\bar{\zeta}_n} \left|\zeta^2 \int_{|x_2|>\bar{x}_n} \left[\mathbf{i}\check{s}_{YX}(x_2) - \mathbf{i}x_2\check{s}_Y(x_2) - (\mathbf{i}\hat{s}_{YX}(x_2) - \mathbf{i}x_2\hat{s}_Y(x_2))\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\le \sup_{|\zeta|\le\bar{\zeta}_n} \left|\zeta^2 \int_{|x_2|>\bar{x}_n} \left[\mathbf{i}\check{s}_{YX}(x_2) - \mathbf{i}x_2\check{s}_Y(x_2) - (\mathbf{i}s_{YX}(x_2) - \mathbf{i}x_2s_Y(x_2))\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$+ \sup_{|\zeta|\le\bar{\zeta}_n} \left|\zeta^2 \int_{|x_2|>\bar{x}_n} \left[\mathbf{i}\hat{s}_{YX}(x_2) - \mathbf{i}x_2\hat{s}_Y(x_2) - (\mathbf{i}s_{YX}(x_2) - \mathbf{i}x_2s_Y(x_2))\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\le \bar{\zeta}_n^2 \int_{|x_2|>\bar{x}_n} |\check{s}_{YX}(x_2) - s_{YX}(x_2)|\, dx_2 + \bar{\zeta}_n^2 \int_{|x_2|>\bar{x}_n} |x_2|\, |\check{s}_Y(x_2) - s_Y(x_2)|\, dx_2$$

$$+ \sup_{|\zeta|\le\bar{\zeta}_n} \left|\int \left[\nabla^2\hat{s}_{YX}(x_2) - \nabla^2 s_{YX}(x_2)\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$+ \sup_{|\zeta|\le\bar{\zeta}_n} \left|\int \left[2\nabla\hat{s}_Y(x_2) - 2\nabla s_Y(x_2)\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$+ \sup_{|\zeta|\le\bar{\zeta}_n} \left|\int \left[x_2\nabla^2\hat{s}_Y(x_2) - x_2\nabla^2 s_Y(x_2)\right] e^{\mathbf{i}\zeta x_2}dx_2\right|.$$

The two terms in the third-to-last line are $O(T_{YX,n}^{d=0,j=0,k=2} + T_{Y,n}^{d=0,j=1,k=2})$ by definition. Splitting the three terms in the last two lines into integrals over $\{|x_2| \le \bar{x}_n\}$ and over $\{|x_2| > \bar{x}_n\}$ yields

$$\sup_{|\zeta|\le\bar{\zeta}_n} \left|\int \left[\nabla^2\hat{s}_{YX}(x_2) - \nabla^2 s_{YX}(x_2)\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\le \underline{f}_n^{-1/2}\|\nabla^2\hat{s}_{YX} - \nabla^2 s_{YX}\| + O(T_{YX,n}^{d=2,j=0,k=0})$$

$$\sup_{|\zeta|\le\bar{\zeta}_n} \left|\int \left[2\nabla\hat{s}_Y(x_2) - 2\nabla s_Y(x_2)\right] e^{\mathbf{i}\zeta x_2}dx_2\right| \le 2\underline{f}_n^{-1/2}\|\nabla\hat{s}_Y - \nabla s_Y\| + O(T_{Y,n}^{d=1,j=0,k=0})$$

$$\sup_{|\zeta|\le\bar{\zeta}_n} \left|\int \left[x_2\nabla^2\hat{s}_Y(x_2) - x_2\nabla^2 s_Y(x_2)\right] e^{\mathbf{i}\zeta x_2}dx_2\right|$$

$$\le \underline{f}_n^{-1/2}\bar{x}_n\|\nabla^2\hat{s}_Y - \nabla^2 s_Y\| + O(T_{Y,n}^{d=2,j=1,k=0}),$$

so, by collecting terms,

$$\sup_{|\zeta|\le\bar\zeta_n} \left|\zeta^2\hat\sigma_\Delta(\zeta) - \zeta^2[\mathcal{F}\hat s_\Delta](\zeta)\right|$$

$$= O_p(\underline{f}_n^{-1/2}(\omega_{2,n} + \omega_{1,n} + \omega_{2,n}\bar x_n)\delta_n)$$
$$+ O(T_{YX,n}^{d=0,j=0,k=2} + T_{Y,n}^{d=0,j=1,k=2} + T_{YX,n}^{d=2,j=0,k=0} + T_{Y,n}^{d=1,j=0,k=0} + T_{Y,n}^{d=2,j=1,k=0})$$
$$= O_p(\underline{f}_n^{-1/2}\omega_{2,n}\bar x_n\delta_n) + O(T_{YX,n}^{d=0,j=0,k=2} + T_{Y,n}^{d=0,j=1,k=2})$$
$$+ O(T_{YX,n}^{d=2,j=0,k=0} + T_{Y,n}^{d=1,j=0,k=0} + T_{Y,n}^{d=2,j=1,k=0}).$$

Similarly,

$$\sup_{|\zeta|\le\bar\zeta_n} \left|\zeta^2[\mathcal{F}\hat s_\Delta](\zeta) - \zeta^2\sigma_\Delta(\zeta)\right|$$

$$= O_p(\underline{f}_n^{-1/2}\omega_{2,n}\bar x_n\delta_n) + O(T_{YX,n}^{d=2,j=0,k=0} + T_{Y,n}^{d=1,j=0,k=0} + T_{Y,n}^{d=2,j=1,k=0})$$

and, thus, letting $T_{\Delta,n} := T_{Y,n}^{0,1,2} + T_{Y,n}^{2,1,0} + T_{Y,n}^{1,0,0} + T_{YX,n}^{0,0,2} + T_{YX,n}^{2,0,0}$, we have

$$\sup_{|\zeta|\le\bar\zeta_n} |\hat\sigma_\Delta(\zeta) - \sigma_\Delta(\zeta)| = O_p(\underline{f}_n^{-1/2}\omega_{2,n}\bar x_n\delta_n\bar\zeta_n^{-2}) + O(T_{\Delta,n}). \tag{43}$$

Step (iii) By Assumption C 4 and $\bar r_n^{-1} = O_p(1)$, we have $\underline{f}_n^{-1/2}\delta_n\omega_{1,n}\bar\zeta^{-1}\underline\sigma_n^{-1} \to 0$ and $T_{Y,n}\underline\sigma_n^{-1} \to 0$. Therefore, Lemma 3 of Schennach (2008) can be applied, and, together with (42) and (43), yields

$$\bar\mu_n := \sup_{|\zeta|\le\bar\zeta_n} \left|\frac{\hat\sigma_\Delta(\zeta)}{\hat\sigma_Y(\zeta)} - \frac{\sigma_\Delta(\zeta)}{\sigma_Y(\zeta)}\right|$$

$$= O_p\left(\left\{\sup_{|\zeta|\le\bar\zeta_n} |\hat\sigma_\Delta(\zeta) - \sigma_\Delta(\zeta)|\right\}\underline\sigma_n^{-1}\right) + O_p\left(\bar r_n\left\{\sup_{|\zeta|\le\bar\zeta_n} |\hat\sigma_Y(\zeta) - \sigma_Y(\zeta)|\right\}\underline\sigma_n^{-1}\right)$$

$$= O_p\left(\left\{\underline{f}_n^{-1/2}\omega_{2,n}\bar x_n\delta_n\bar\zeta_n^{-2} + T_{\Delta,n}\right\}\underline\sigma_n^{-1}\right) + O_p\left(\bar r_n\left\{\underline{f}_n^{-1/2}\delta_n\omega_{1,n}\bar\zeta_n^{-1} + T_{Y,n}\right\}\underline\sigma_n^{-1}\right)$$

$$= O_p\left(\underline{f}_n^{-1/2}\left\{\omega_{2,n}\bar x_n\delta_n\bar\zeta_n^{-2} + \bar r_n\delta_n\omega_{1,n}\bar\zeta_n^{-1}\right\}\underline\sigma_n^{-1}\right) + O_p\left((\bar r_n T_{Y,n} + T_{\Delta,n})\underline\sigma_n^{-1}\right).$$

The last two equations use the convergence rates from step (ii).

Step (iv) This step is inspired in part by the proof of Theorem 2 in Schennach (2008). Decompose the estimation error into three parts:

$$2\pi\left|\hat F_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*)\right|$$

$$= \left|\int_{|\zeta|\le\bar\zeta_n} \hat\sigma_Y(\zeta,y_2)\hat\phi(\zeta,y_2)e^{-\mathbf{i}\zeta\cdot x_2^*}d\zeta - \int \sigma_Y(\zeta,y_2)\phi(\zeta,y_2)e^{-\mathbf{i}\zeta\cdot x_2^*}d\zeta\right|$$

$$\le R_1 + R_2 + R_3,$$

where

$$R_1 := \left| \int_{|\zeta| \leq \bar{\zeta}_n} \hat{\sigma}_Y(\zeta, y_2) \left[\hat{\phi}(\zeta, y_2) - \phi(\zeta, y_2) \right] e^{-i\zeta \cdot x_2^*} d\zeta \right|$$

$$R_2 := \left| \int_{|\zeta| \leq \bar{\zeta}_n} \left[\hat{\sigma}_Y(\zeta, y_2) - \sigma_Y(\zeta, y_2) \right] \phi(\zeta, y_2) e^{-i\zeta \cdot x_2^*} d\zeta \right|$$

$$R_3 := \left| \int_{|\zeta| > \bar{\zeta}_n} \sigma_Y(\zeta, y_2) \phi(\zeta, y_2) e^{-i\zeta \cdot x_2^*} d\zeta \right|.$$

First, notice $\sigma_Y(\zeta, y_2) = \gamma(\zeta, y_2)/\phi(\zeta)$ diverges not only at the origin, but potentially also as $|\zeta| \to \pm\infty$ when ϕ vanishes faster than γ. For this reason, I split up the first remainder R_1 further into R_{1a} and R_{1b}, which, respectively, bound the error for small $|\zeta|$ around the origin and for those up to $\bar{\zeta}_n$. Formally, fix some constant $\zeta_0 \in (0, \bar{\zeta}_n)$ and write $R_1 \leq R_{1a} + R_{1b}$ with

$$R_{1a} := \left| \int_{|\zeta| \leq \zeta_0} \hat{\sigma}_Y(\zeta, y_2) \left[\hat{\phi}(\zeta, y_2) - \phi(\zeta, y_2) \right] e^{-i\zeta \cdot x_2^*} d\zeta \right|,$$

$$R_{1b} := \left| \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} \hat{\sigma}_Y(\zeta, y_2) \left[\hat{\phi}(\zeta, y_2) - \phi(\zeta, y_2) \right] e^{-i\zeta \cdot x_2^*} d\zeta \right|,$$

Consider the first remainder. On the interval $(0, \zeta_0)$, Lemma 4(ii) bounds $|\sigma_Y(\zeta, y_2)|$ from above by $\bar{\sigma} \max\{|\zeta|^{-1}, 1\}$ for some constant $\bar{\sigma}$. Using this fact, we have

$$R_{1a} = \left| \int_{|\zeta| \leq \zeta_0} \hat{\sigma}_Y(\zeta, y_2) \left[\exp\left\{ \int_0^\zeta \frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} dz \right\} - \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right] e^{-i\zeta \cdot x_2^*} d\zeta \right|$$

$$\leq \int_{|\zeta| \leq \zeta_0} |\hat{\sigma}_Y(\zeta, y_2)| \left| \exp\left\{ \int_0^\zeta \frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} dz \right\} - \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right| d\zeta$$

$$= \int_{|\zeta| \leq \zeta_0} |\sigma_Y(\zeta, y_2) + \epsilon_n| \left| \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right| \times$$

$$\times \left| \exp\left\{ \int_0^\zeta \frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} dz - \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} - 1 \right| d\zeta$$

$$= C \int_{|\zeta| \leq \zeta_0} |\sigma_Y(\zeta, y_2) + \epsilon_n| \left| \exp\left\{ \int_0^\zeta \left(\frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} - \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} \right) dz \right\} - 1 \right| d\zeta,$$

where $\epsilon_n = O_p(\underline{f}_n^{-1/2} \delta_n \omega_{1,n} \bar{\zeta}_n^{-1}) + O(T_{Y,n})$, which occurs from the use of the rate in (42). Step (iii) gives a bound for the difference of ratios such that

$$R_{1a} \leq C \int_{|\zeta| \leq \zeta_0} \left(\bar{\sigma} \max\{|\zeta|^{-1}, 1\} + \epsilon_n \right) \left| e^{\bar{\mu}_n \zeta} - 1 \right| d\zeta$$

$$\leq C \int_{|\zeta| \leq \zeta_0} \left(\bar{\sigma} \max\{|\zeta|^{-1}, 1\} + \epsilon_n \right) |\bar{\mu}_n \zeta| d\zeta$$

46

$$\leq C \int_{|\zeta| \leq \zeta_0} \left(\bar{\sigma} \max\{1, |\zeta|\} + |\zeta| \epsilon_n \right) \bar{\mu}_n d\zeta$$

$$\leq 2C \left(\bar{\sigma} \max\{1, \zeta_0\} + \zeta_0 \epsilon_n \right) \zeta_0 \bar{\mu}_n$$

$$= O_p(\epsilon_n \bar{\mu}_n + \bar{\mu}_n),$$

where the second inequality uses the series expansion of the exponential function. The remainder R_{1b} is treated in a similar way:

$$R_{1b} = \left| \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} \hat{\sigma}_Y(\zeta, y_2) \left[\exp\left\{ \int_0^\zeta \frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} dz \right\} - \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right] e^{-i\zeta \cdot x_2^*} d\zeta \right|$$

$$\leq \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\hat{\sigma}_Y(\zeta, y_2)| \left| \exp\left\{ \int_0^\zeta \frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} dz \right\} - \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right| d\zeta$$

$$= \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\sigma_Y(\zeta, y_2) + \epsilon_n| \left| \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right| \times$$

$$\times \left| \exp\left\{ \int_0^\zeta \frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} dz - \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} - 1 \right| d\zeta$$

$$= \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\sigma_Y(\zeta, y_2) + \epsilon_n| \, |\phi(\zeta)| \left| \exp\left\{ \int_0^\zeta \left(\frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} - \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} \right) dz \right\} - 1 \right| d\zeta$$

$$\leq \left(O(1) + \epsilon_n \sup_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\phi(\zeta)| \right) \times$$

$$\times \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} \left| \exp\left\{ \int_0^\zeta \left(\frac{\hat{\sigma}_\Delta(z, y_2)}{\hat{\sigma}_Y(z)} - \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} \right) dz \right\} - 1 \right| d\zeta$$

$$\leq \left(O(1) + \epsilon_n \sup_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\phi(\zeta)| \right) \int_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\bar{\mu}_n \zeta| d\zeta$$

$$\leq \left(O(1) + \epsilon_n \sup_{\zeta_0 < |\zeta| \leq \bar{\zeta}_n} |\phi(\zeta)| \right) 2\bar{\mu}_n \left(\bar{\zeta}_n^2 + C'' \right)$$

$$\leq C''' \bar{\mu}_n \bar{\zeta}_n^2 + 2\epsilon_n \bar{\mu}_n \bar{\zeta}_n^2 + C^{(4)} \bar{\mu}_n + C^{(5)} \epsilon_n \bar{\mu}_n,$$

where the third inequality uses step (iii) as before. The second remainder can be bounded as follows:

$$R_2 = \left| \int_{|\zeta| \leq \bar{\zeta}_n} [\hat{\sigma}_Y(\zeta, y_2) - \sigma_Y(\zeta, y_2)] \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} e^{-i\zeta \cdot x_2^*} d\zeta \right|$$

$$\leq \int_{|\zeta| \leq \bar{\zeta}_n} \left| [\hat{\sigma}_Y(\zeta, y_2) - \sigma_Y(\zeta, y_2)] \exp\left\{ \int_0^\zeta \frac{\sigma_\Delta(z, y_2)}{\sigma_Y(z)} dz \right\} \right| d\zeta$$

$$\leq \left[\sup_{|\zeta| \leq \bar{\zeta}_n} \phi(\zeta) \right] \int_{|\zeta| \leq \bar{\zeta}_n} |\hat{\sigma}_Y(\zeta, y_2) - \sigma_Y(\zeta, y_2)| \, d\zeta$$

47

$$\leq 2\bar{\zeta}_n \epsilon_n,$$

where the last inequality follows from step (ii). Finally, the last term R_3 represents a tail-trimming error:

$$R_3 = \left| \int_{|\zeta| > \bar{\zeta}_n} \gamma(\zeta, y_2) e^{-i\zeta \cdot x_2^*} d\zeta \right| \leq \int_{|\zeta| > \bar{\zeta}_n} |\gamma(\zeta, y_2)| d\zeta = O(T_{\gamma, n}).$$

Now, combine the three remainders to get

$$2\pi \sup_{x_2^* \in \mathbb{R}} \left| \hat{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right|$$

$$= O_p\left(\bar{\mu}_n + \epsilon_n \bar{\mu}_n + \bar{\mu}_n \bar{\zeta}_n^2 + \epsilon_n \bar{\mu}_n \bar{\zeta}_n^2 + \bar{\mu}_n + \epsilon_n \bar{\mu}_n \right) + O_p\left(\bar{\zeta}_n \epsilon_n \right) + O(T_{\gamma, n})$$

$$= O_p\left(\bar{\mu}_n \bar{\zeta}_n^2 + \bar{\zeta}_n \epsilon_n \right) + O(T_{\gamma, n}).$$

The first equality holds for the following reason. The conditions $\bar{r}_n \delta_n \omega_{1,n} / (\underline{f}_n^{1/2} \bar{\zeta}_n \underline{\sigma}_n) \to 0$ and $\bar{r}_n T_{Y,n} / \underline{\sigma}_n \to 0$ in Assumption C 4 imply $\bar{\mu}_n \to 0$. In addition, $\bar{r}_n^{-1} = O(1)$ and $\underline{\sigma}_n = O(1)$, so $\delta_n \omega_{1,n} / (\underline{f}_n^{1/2} \bar{\zeta}_n) \to 0$ and $T_{Y,n} \to 0$, leading to $\epsilon_n \to 0$. Next, substituting in the rates of $\bar{\mu}_n$ and ϵ_n yields

$$2\pi \sup_{x_2^* \in \mathbb{R}} \left| \hat{F}_{Y_2|X_2^*}(y_2|x_2^*) - F_{Y_2|X_2^*}(y_2|x_2^*) \right|$$

$$= O_p\left(\bar{\zeta}_n^2 \left[\underline{f}_n^{-1/2} \left\{ \omega_{2,n} \bar{x}_n \delta_n \bar{\zeta}_n^{-2} + \bar{r}_n \delta_n \omega_{1,n} \bar{\zeta}_n^{-1} \right\} \underline{\sigma}_n^{-1} + (\bar{r}_n T_{Y,n} + T_{\Delta,n}) \underline{\sigma}_n^{-1} \right] \right)$$

$$\quad + O_p\left(\bar{\zeta}_n \left[\underline{f}_n^{-1/2} \delta_n \omega_{1,n} \bar{\zeta}_n^{-1} + T_{Y,n} \right] \right) + O(T_{\gamma, n})$$

$$= O_p\left(\underline{f}_n^{-1/2} \left\{ \omega_{2,n} \bar{x}_n \delta_n + \bar{r}_n \delta_n \omega_{1,n} \bar{\zeta}_n \right\} \underline{\sigma}_n^{-1} + T_n \right), \tag{44}$$

which gives the convergence rate pointwise in y_2. The expression in (44) is also an upper bound on the convergence rate of the constraint estimator $\check{F}_{Y_2|X_2^*}(y_2|x_2^*)$. Since $\check{F}_{Y_2|X_2^*}(y_2|x_2^*)$ takes values only in $[0, 1]$, the convergence rate holds, in fact, uniformly over $y_2 \in \mathbb{Y}_2$. This conclusion follows by essentially the same argument as the proof of the Glivenko-Cantelli Theorem; see Theorem 19.1 of van der Vaart (1998), for example. This completes the proof.   Q.E.D.

Proof of Corollary 3 First, I establish the order of the tail-trimming term T_n. To that end, let $\gamma_n := \sup_{y_2 \in \mathbb{Y}} |\gamma(\bar{\zeta}_n, y_2)|$ and $s_n := \sup_{y_2 \in \mathbb{Y}} |\nabla s_Y(\bar{x}_n, y_2)|$.

Consider the term $T_{Y,n}$. Analogously to the proof of Lemma 1 in Schennach (2008), we have $\lim_{x_2 \to -\infty} s_Y(x_2, y_2) = c_Y^-(y_2)$. Therefore, using the definition of $\hat{s}_Y(x_2, y_2)$ when

48

$x_2 < -\bar{x}_n$ and the Fundamental Theorem of Calculus, one can write

$$\int_{-\infty}^{-\bar{x}_n} |x_2|^j \left| -c_Y^-(y_2) - s_Y(x_2, y_2) \right| dx_2 = \int_{-\infty}^{-\bar{x}_n} |x_2|^j \left| -\int_{-\infty}^{x_2} \frac{\partial s_Y(u, y_2)}{\partial x_2} du \right| dx_2$$

$$\leq C_s \int_{-\infty}^{-\bar{x}_n} |x_2|^j \int_{-\infty}^{x_2} (1 + |u|)^{\gamma_s} \exp\left\{ -\alpha_s |u|^{\beta_s} \right\} du dx_2$$

$$= C_s \int_{\bar{x}_n}^{\infty} x_2^j \int_{x_2}^{\infty} (1 + u)^{\gamma_s} \exp\left\{ -\alpha_s u^{\beta_s} \right\} du dx_2$$

$$= C_s \int_{\bar{x}_n}^{\infty} x_2^j (1 + x_2)^{\gamma_s - \beta_s + 1} \exp\left\{ -\alpha_s x_2^{\beta_s} \right\} dx_2$$

$$= O\left((1 + \bar{x}_n)^{\gamma_s - 2\beta_s + 2 + j} \exp\left\{ -\alpha_s \bar{x}_n^{\beta_s} \right\} \right)$$

$$= O\left(\bar{x}_n^{-2\beta_s + 2 + j} s_n \right) \tag{45}$$

by Assumption R 3.2 and repeated application of Lemma 4.2 in Li and Vuong (1998). In the same fashion, one can show that $\int_{\bar{x}_n}^{\infty} |x_2|^j |c_Y^+(y_2) - s_Y(x_2, y_2)| dx_2 = O(\bar{x}_n^{-2\beta_s + 2 + j} s_n)$ as well. Therefore, $T_{Y,n}^{0,j,1} = O(\bar{\zeta}_n \bar{x}_n^{-2\beta_s + 2 + j} s_n)$ and we have

$$T_{Y,n} = O(T_{Y,n}^{0,j=0,1}) = O(\bar{\zeta}_n \bar{x}_n^{-2\beta_s + 2} s_n).$$

Next, consider the second component, $T_{\Delta,n}$. Observe

$$\int_{-\infty}^{-\bar{x}_n} \left| x_2 c_Y^-(y_2) - s_{YX}(x_2, y_2) \right| dx_2 = \int_{-\infty}^{-\bar{x}_n} |x_2| \left| c_Y^-(y_2) - \frac{s_{YX}(x_2, y_2)}{x_2} \right| dx_2$$

$$\asymp \int_{-\infty}^{-\bar{x}_n} |x_2| \left| c_Y^-(y_2) - s_Y(x_2, y_2) \right| dx_2 \tag{46}$$

$$= O\left(\bar{x}_n^{-2\beta_s + 3} s_n \right), \tag{47}$$

where the last equality uses (45). The asymptotic equality in (46) can be justified as follows. From the two equations (27) and (28), we have that

$$E[s_{YX}(X_2, y_2) | X_2^* = x_2^*] = x_2^* E[s_Y(X_2, y_2) | X_2^* = x_2^*] = E[X_2 s_Y(X_2, y_2) | X_2^* = x_2^*]$$

for all $y_2 \in \mathbb{Y}_2$. Since $\mathcal{C}$ is injective, the distribution of X_2 given X_2^* is complete; that is, $E[s_{YX}(X_2, y_2) - X_2 s_Y(X_2, y_2) | X_2^* = x_2^*] = 0$ implies $s_{YX}(X_2, y_2) = X_2 s_Y(X_2, y_2) P_{X_2}$-almost surely.

Similarly as in (47), we also have

$$\int_{-\infty}^{-\bar{x}_n} \left| x_2 c_Y^-(y_2) - s_{YX}(x_2, y_2) \right| dx_2 = O\left(\bar{x}_n^{-2\beta_s + 4} s_n \right). \tag{48}$$

By similar reasoning as for the remainder $T_{Y,n}$, it is easy to see that $T_{\Delta,n} = O(T_{Y,n}^{0,1,2} + T_{YX,n}^{0,0,2})$. Therefore, (48) and

49

Next, using again Lemma 4.2 in Li and Vuong (1998), it is easy to see the third component of T_n, $\int_{|\zeta|>\bar\zeta_n}|\gamma(\zeta,y_2)|d\zeta$, is of order $O(\bar\zeta_n^{-\beta_g+1}\gamma_n)$. In conclusion, we have $T_n = O(\bar x_n^{-2\beta_s+3}s_n + \bar\zeta_n^{-\beta_g+1}\gamma_n)$.

Now consider

$$
\beta_n := \bar\zeta_n^2 \bar r_n \left[\underline{f}_n^{-1/2}\delta_n + \bar x_n^{-2\beta_s+3}s_n + \bar\zeta_n^{-\beta_g+1}\gamma_n\right]\underline\sigma_n^{-1}
$$

$$
= O\left(\bar\zeta_n^{2+\gamma_r-\gamma_g}e^{\alpha_g\bar\zeta_n^{\beta_g}}\bar x_n^{-\gamma_{fX}/2}e^{\alpha_{fX}\bar x_n^{\beta_{fX}}/2}\delta_n + \bar\zeta_n^{2+\gamma_r-\gamma_g}e^{\alpha_g\bar\zeta_n^{\beta_g}}\bar x_n^{\gamma_s-2\beta_s+3}e^{-\alpha_s\bar x_n^{\beta_s}}\right.
$$

$$
\left. +\bar\zeta_n^{2+\gamma_r-\gamma_g}e^{\alpha_g\bar\zeta_n^{\beta_g}}\bar\zeta_n^{\gamma_g-\beta_g+1}e^{-\alpha_g\bar\zeta_n^{\beta_g}}\right)
$$

$$
= O(\bar\zeta_n^{2+\gamma_r-\gamma_g}\bar x_n^{-\gamma_{fX}/2}e^{\alpha_g\bar\zeta_n^{\beta_g}+\alpha_{fX}\bar x_n^{\beta_{fX}}/2}\delta_n + \bar\zeta_n^{2+\gamma_r-\gamma_g}\bar x_n^{\gamma_s-2\beta_s+3}e^{\alpha_g\bar\zeta_n^{\beta_g}-\alpha_s\bar x_n^{\beta_s}} + \bar\zeta_n^{\gamma_r-\beta_g+3}).
$$

$$(49)$$

To balance the bias and variance terms in $K_n^{-\rho_s}+K_n^\omega\sqrt{K_n/n}$, we select $K_n = O(n^{1/[2(\rho_s+\omega)+1]})$ and get $\delta_n = O(n^{-\rho_s/[2(\rho_s+\omega)+1]})$.

The remainder of the proof consists in merely substituting in the given expressions for $\bar x_n$ and $\bar\zeta_n$, and checking that the rates are $o_p(1)$ under the stated assumptions on the various parameters. Q.E.D.

Proof of Corollary 4   In the severely ill-posed case, we select $K_n = \log(n)$ to balance the bias and variance terms in $\delta_n = K_n^{-\rho_s}+\exp\{K_n\}\sqrt{K_n/n}$. Similarly as in the derivation of (49),

$$
\beta_n = O(\bar\zeta_n^{2+\gamma_r-\gamma_g}\bar x_n^{-\gamma_{fX}/2}e^{-\alpha_r\bar\zeta_n^{\beta_r}+\alpha_g\bar\zeta_n^{\beta_g}+\alpha_{fX}\bar x_n^{\beta_{fX}}/2}\delta_n
$$

$$
+ \bar\zeta_n^{2+\gamma_r-\gamma_g}\bar x_n^{\gamma_s-2\beta_s+3}e^{-\alpha_r\bar\zeta_n^{\beta_r}+\alpha_g\bar\zeta_n^{\beta_g}-\alpha_s\bar x_n^{\beta_s}} + \bar\zeta_n^{\gamma_r-\beta_g+3}e^{-\alpha_r\bar\zeta_n^{\beta_r}}). \quad (50)
$$

As in the proof of the previous corollary, the remainder of this proof consists of merely substituting in the given expressions for $\bar x_n$ and $\bar\zeta_n$, and checking that the rates are $o_p(1)$ under the stated assumptions on the various parameters. Q.E.D.

Proof of Theorem 4   For the quantile estimation case, see the proof of Theorem 3.1 in Ould-Saïd, Yahia, and Necir (2009). If the conditional mean restriction holds, then integration by parts and the fact that the rates in Corollaries 3 or 4 are uniform over $\mathbb{Y}_2$ yield the desired result. Q.E.D.

References

AASNESS, J., E. BIØRN, AND T. SKJERPEN (1993): "Engel Functions, Panel Data, and Latent Variables," *Econometrica*, 61(6), 1395–1422.

ALMEIDA, H., M. CAMPELLO, AND A. F. GALVAO (2010): "Measurement Errors in Investment Equations," *Review of Financial Studies*, 23(9), 3279–3328.

ANDREWS, D. W. K. (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11(3), 560–596.

——— (2011): "Examples of L^2-Complete and Boundedly-Complete Distributions," Discussion Paper 1801, Cowles Foundation, Yale University.

ARELLANO, M. (2014): "Uncertainty, Persistence, and Heterogeneity: A Panel Data Perspective," *Journal of the European Economic Association*, 12(5), 1127–1153.

ARELLANO, M., R. BLUNDELL, AND S. BONHOMME (2014): "Household Earnings and Consumption: A Nonlinear Framework," Discussion paper.

BARNETT, S. A., AND P. SAKELLARIS (1998): "Nonlinear Response of Firm Investment to Q: Testing a Model of Convex and Non-Convex Adjustment Costs," *Journal of Monetary Economics*, 42(2), 261–288.

BERTRAND, M., AND A. SCHOAR (2003): "Managing with Style: The Effect of Managers on Firm Policies," *The Quarterly Journal of Economics*, 118(4), 1169–1208.

BIØRN, E. (2000): "Panel Data With Measurement Errors: Instrumental Variables And Gmm Procedures Combining Levels And Differences," *Econometric Reviews*, 19(4), 391–424.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613–1669.

BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. V, pp. 3705–3843. Elsevier Science B.V.

BOUND, J., AND A. B. KRUEGER (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics*, 9(1), 1–24.

BUONACCORSI, J., E. DEMIDENKO, AND T. TOSTESON (2000): "Estimation in Longitudinal Random Effects Models with Measurement Error," *Statistica Sinica*, 10, 885–903.

CANAY, I. A., A. SANTOS, AND A. M. SHAIKH (2013): "On the Testability of Identification in Some Nonparametric Models With Endogeneity," *Econometrica*, 81(6), 2535–2559.

CARD, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64(4), 957–979.

CARRASCO, M., AND J.-P. FLORENS (2011): "A Spectral Method for Deconvolving a Density," *Econometric Theory*, 7(Special Issue 03), 546–581.

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. VI, pp. 5633–5751. Elsevier Science B.V.

CARROLL, R. J., A. C. M. ROOIJ, AND F. H. RUYMGAART (1991): "Theoretical Aspects of Ill-posed Problems in Statistics," *Acta Applicandae Mathematicae*, 24(2), 113–140.

CARROLL, R. J., D. RUPPERT, L. A. STEFANSKI, AND C. M. CRAINICEANU (2006): *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall, New York.

CHEN, X., H. HONG, AND D. NEKIPELOV (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49(4), 901–937.

CHEN, X., AND M. REISS (2011): "On Rate Optimality for Ill-Posed Inverse Problems in Econometrics," *Econometric Theory*, 27(Special Issue 03), 497–521.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND A. GALICHON (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78(3), 1093–1125.

CHESHER, A. (1991): "The Effect of Measurement Error," *Biometrika*, 78(3), 451–462.

CHETVERIKOV, D., AND D. WILHELM (2015): "Nonparametric Instrumental Variable Estimation Under Monotonicity," Discussion paper.

CRAGG, J. G., AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9(2), 222–240.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

DAROLLES, S., Y. FAN, J. P. FLORENS, AND E. RENAULT (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79(5), 1541–1565.

D'HAULTFŒUILLE, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 7(Special Issue 03), 460–471.

ERICKSON, T., AND T. M. WHITED (2000): "Measurement Error and the Relationship between Investment and q," *Journal of Political Economy*, 108(5), 1027–1057.

——— (2012): "Treating Measurement Error in Tobin's q," *Review of Financial Studies*, 25(4), 1286–1329.

EVDOKIMOV, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Discussion paper, Princeton University.

FAN, J. (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19(3), 1257–1272.

FAN, J., AND Y. K. TRUONG (1993): "Nonparametric Regression with Errors in Variables," *The Annals of Statistics*, 21(4), 1900–1925.

FÈVE, F., AND J.-P. FLORENS (2014): "Non parametric analysis of panel data models with endogenous variables," *Journal of Econometrics*, 181(2), 151–164.

GALVAO JR., A. F., AND G. MONTES-ROJAS (2009): "Instrumental Variables Quantile Regression for Panel Data with Measurement Errors," Discussion paper, City University London.

GRILICHES, Z., AND J. A. HAUSMAN (1986): "Errors in Variables in Panel Data," *Journal of Econometrics*, 31, 93–118.

HALL, B. H., A. B. JAFFE, AND M. TRAJTENBERG (2005): "Market Value and Patent Citations," *The RAND Journal of Economics*, 36(1), 16–38.

HALL, P., AND J. L. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33(6), 2904–2929.

HAUSMAN, J., W. NEWEY, H. ICHIMURA, AND J. POWELL (1991): "Identification and Estimation of Polynomial Errors-in-Variables Models," *Journal of Econometrics*, 50, 273–295.

HAYASHI, F. (1982): "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50(1), 213–224.

HENDERSON, D. J., R. J. CARROLL, AND Q. LI (2008): "Nonparametric Estimation and Testing of Fixed Effects Panel Data Models," *Journal of Econometrics*, 144(1), 257–275.

HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371–1395.

HOROWITZ, J., AND S. LEE (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75(4), 1191–1208.

HOROWITZ, J. L. (2009): "Specification Testing in Nonparametric Instrumental Variables Estimation," Discussion paper, Northwestern University.

——— (2011): "Applied Nonparametric Instrumental Variables Estimation," *Econometrica*, 79(2), 347–394.

HU, Y., AND S. M. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

HU, Y., AND J.-L. SHIU (2011): "Nonparametric Identification Using Instrumental Variables: Sufficient Conditions For Completeness," Discussion paper, Johns Hopkins University.

JOHANNES, J., S. VAN BELLEGEM, AND A. VANHEMS (2011): "Convergence Rates for Ill-Posed Inverse Problems with an Unknown Operator," *Econometric Theory*, 7(Special Issue 03), 522–545.

KHAN, S., M. PONOMAREVA, AND E. TAMER (2011): "Identification of Panel Data Models with Endogenous Censoring," Discussion paper, Duke University.

KOMUNJER, I., AND S. NG (2011): "Measurement Errors in Dynamic Models," *Econometric Theory*, forthcoming.

KOOPMANS, T. C., AND O. REIERSOL (1950): "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21(2), 165–181.

KOTLARSKI, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20(1), 69–76.

LEE, Y. (2014): "Nonparametric Estimation of Dynamic Panel Models with Fixed Effects," *Econometric Theory*, 30, 1315–1347.

LI, T., AND Q. VUONG (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65(2), 139–165.

LUCAS, JR., R. E., AND E. C. PRESCOTT (1971): "Investment Under Uncertainty," *Econometrica*, 39(5), 659–681.

MATZKIN, R. L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, pp. 2523–2558. Elsevier Science B.V.

——— (2007): "Nonparametric Identification," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. VI, pp. 5307–5368. Elsevier Science B.V.

NEWEY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147–168.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

OLLEY, G. S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

OULD-SAÏD, E., D. YAHIA, AND A. NECIR (2009): "A Strong Uniform Convergence Rate of a Kernel Conditional Quantile Estimator under Random Left-Truncation and Dependent Data," *Electronic Journal of Statistics*, 3, 426–445.

PORTER, J. R. (1996): "Essays in Econometrics," Ph.D. thesis, MIT.

QIAN, J., AND L. WANG (2012): "Estimating Semiparametric Panel Data Models by Marginal Integration," *Journal of Econometrics*, 167(2), 483–493.

ROEHRIG, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56(2), 433–447.

SCHENNACH, S. M. (2004a): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72(1), 33–75.

——— (2004b): "Nonparametric Regression in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.

———— (2007): "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75(1), 201–239.

———— (2008): "Quantile Regression with Mismeasured Covariates," *Econometric Theory*, 24, 1010–1043.

———— (2013): "Measurement Error in Nonlinear Models — A Review," in *Advances in Economics and Econometrics*, vol. 3. Cambridge University Press.

SCHENNACH, S. M., AND Y. HU (2013): "Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models Without Side Information," *Journal of the American Statistical Association*, 108(501), 177–186.

SHAO, J., Z. XIAO, AND R. XU (2011): "Estimation with Unbalanced Panel Data Having Covariate Measurement Error," *Journal of Statistical Planning and Inference*, 141(2), 800–808.

STOCK, J. H., AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock, pp. 80–108, Cambridge, UK. Cambridge University Press.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, New York.

WANSBEEK, T. (2001): "GMM Estimation in Panel Data Models with Measurement Error," *Journal of Econometrics*, 104(2), 259–268.

XIAO, Z., J. SHAO, AND M. PALTA (2008): "A Unified Theory for GMM Estimation in Panel Data Models with Measurement Error," Discussion paper, University of Wisconsin-Madison.

		weak ME ($\sigma_\eta = 0.5$)								
		$\alpha_n = 0.001$			$\alpha_n = 0.01$			$\alpha_n = 0.1$		
basis (x_1)		bias	SD	RMSE	bias	SD	RMSE	bias	SD	RMSE
poly	$K_n = 3$	0.054	0.140	0.158	0.054	0.140	0.158	0.066	0.140	0.162
	$K_n = 5$	0.066	0.140	0.162	0.066	0.140	0.162	0.067	0.139	0.162
	$K_n = 7$	0.067	0.137	0.160	0.067	0.137	0.160	0.067	0.137	0.160
spline	$L_n = 5$	0.066	0.141	0.163	0.066	0.141	0.163	0.067	0.140	0.163
	$L_n = 10$	0.068	0.140	0.162	0.068	0.140	0.162	0.068	0.140	0.163
	$L_n = 15$	0.068	0.140	0.162	0.068	0.139	0.162	0.068	0.140	0.163
B-spline	$L_n = 5$	0.071	0.136	0.160	0.071	0.136	0.161	0.072	0.137	0.162
	$L_n = 10$	0.078	0.134	0.162	0.078	0.134	0.163	0.078	0.134	0.163
	$L_n = 15$	0.079	0.139	0.168	0.079	0.139	0.168	0.080	0.138	0.168
NW		0.070	0.125	0.147						

Table 1: For different combinations of the basis in x_1 and the regularization parameter α_n, the table shows average bias, standard deviation (SD) and root mean squared error (RMSE) of the new ME-robust estimator as well as of the Nadaraya-Watson (NW) estimator which ignores the ME.

| basis (x_1) | | strong ME ($\sigma_\eta = 1.5$) | | | | | | | | |
| | | $\alpha_n = 0.001$ | | | $\alpha_n = 0.01$ | | | $\alpha_n = 0.1$ | | |
		bias	SD	RMSE	bias	SD	RMSE	bias	SD	RMSE
poly	$K_n = 3$	0.052	0.328	0.334	0.054	0.326	0.332	0.073	0.313	0.322
	$K_n = 5$	0.051	0.243	0.250	0.051	0.242	0.250	0.057	0.238	0.248
	$K_n = 7$	0.053	0.251	0.258	0.053	0.251	0.258	0.059	0.246	0.256
spline	$L_n = 5$	0.052	0.233	0.242	0.053	0.232	0.242	0.057	0.230	0.241
	$L_n = 10$	0.049	0.221	0.231	0.049	0.221	0.232	0.053	0.219	0.231
	$L_n = 15$	0.049	0.221	0.231	0.049	0.221	0.231	0.053	0.219	0.231
B-spline	$L_n = 5$	0.051	0.217	0.229	0.052	0.217	0.229	0.055	0.216	0.230
	$L_n = 10$	0.065	0.197	0.217	0.065	0.197	0.218	0.068	0.196	0.218
	$L_n = 15$	0.075	0.192	0.218	0.075	0.192	0.218	0.077	0.192	0.219
NW		0.171	0.110	0.207						

Table 2: For different combinations of the basis in x_1 and the regularization parameter α_n, the table shows average bias, standard deviation (SD) and root mean squared error (RMSE) of the new ME-robust estimator as well as of the Nadaraya-Watson (NW) estimator which ignores the ME.
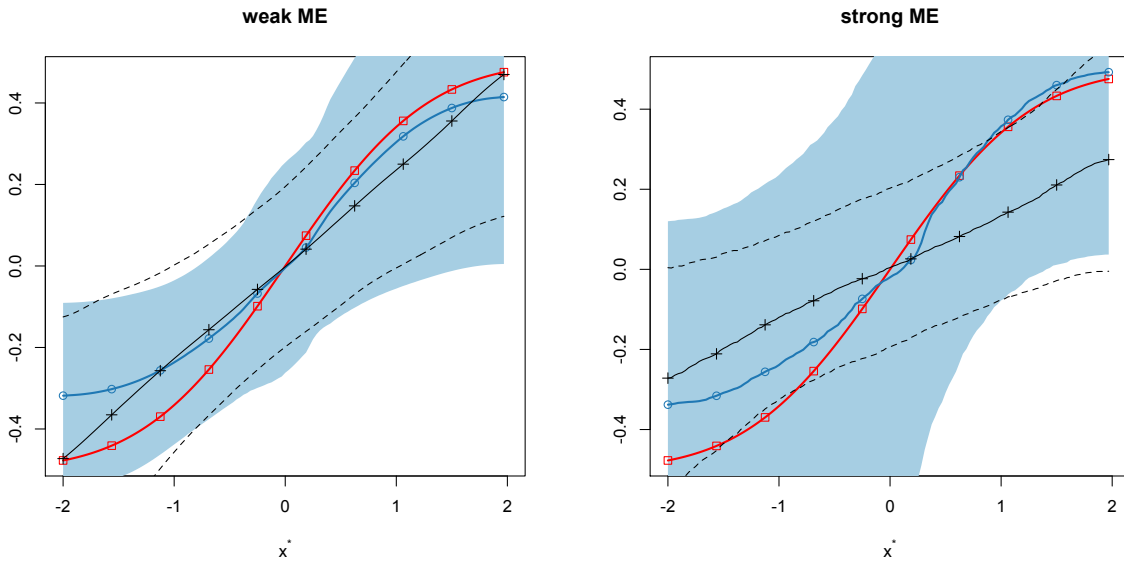

Figure 1: The figure shows the true regression function (red squares), the ME-robust estimator ($\alpha_n = 0.01$, polynomial bases with $K_n = 5$; blue circles) with two empirical standard deviations (shaded area), and the Nadaraya-Watson estimator (black stars) with two empirical standard deviations (dashed line).

basis (x_1)		Cragg-Donald statistic			
		weak ME	strong ME	CV bias	CV size
poly	$K_n = 3$	20.722	33.580	–	8.18
	$K_n = 5$	15.213	27.515	8.78	11.22
	$K_n = 7$	11.312	21.982	9.92	13.34
spline	$L_n = 5$	10.318	21.788	10.43	15.24
	$L_n = 10$	6.444	13.671	10.89	19.72
	$L_n = 15$	4.779	10.140	11.02	24.09
B-spline	$L_n = 5$	10.135	21.472	10.43	15.24
	$L_n = 10$	6.444	13.671	10.89	19.72
	$L_n = 15$	4.779	10.140	11.02	24.09

Table 3: Cragg-Donald statistic for testing the null of weak instruments; large values lead to rejection. Critical values (CV) for 5%-tests of 10% TSLS bias and 15% TSLS size distortion are provided in the last two columns.