Econometric Theory, **38**, 2022, 301–338.
doi:10.1017/S0266466621000116

SEMIPARAMETRIC IDENTIFICATION AND FISHER INFORMATION

JUAN CARLOS ESCANCIANO
Universidad Carlos III de Madrid

This paper provides a systematic approach to semiparametric identification that is based on statistical information as a measure of its "quality." Identification can be regular or irregular, depending on whether the Fisher information for the parameter is positive or zero, respectively. I first characterize these cases in models with densities linear in an infinite-dimensional parameter. I then introduce a novel "generalized Fisher information." If positive, it implies (possibly irregular) identification when other conditions hold. If zero, it implies impossibility results on rates of estimation. Three examples illustrate the applicability of the general results. First, I consider the canonical example of average densities. Second, I show irregular identification of the median willingness to pay in contingent valuation studies. Finally, I study identification of the discount factor and average measures of risk aversion in a nonparametric Euler equation with nonparametric measurement error in consumption.

1. INTRODUCTION

Nonparametric identification is considered the benchmark for reliable empirical analysis in economics. Unfortunately, many nonparametric economic models of interest are unidentified with weak assumptions; see, for example, discrete choice models with nonparametric unobserved heterogeneity (e.g., fixed effects Probit, cf. Chamberlain, 2010). Yet, certain interesting aspects of such models might be point-identified by the same weak assumptions, a situation henceforth referred to as *semiparametric identification*. Although this observation has long been recognized in economics (see the early discussion in Hurwicz, 1950), no systematic method is currently available for assessing which functionals of a nonparametric structural parameter are identified and which functionals are not. Furthermore, even when nonparametric point-identification holds, there could be many parameters that are only "irregularly identified" (in the sense of being identified, but having zero Fisher information; see, e.g., Chamberlain, 1986; Heckman, 1990; Khan and Tamer,

First version: September 20, 2016. Research was funded by the Spanish Programa de Generación de Conocimiento, reference number PGC2018-096732-B-I00. I would like to thank Michael Jansson, Ulrich Müller, Whitney Newey, P.C.B Phillips, Jack Porter, Pedro Sant'Anna, Ruli Xiao, two anonymous referees and seminar participants at BC, Indiana, MIT, Texas A&M, UBC, Vanderbilt, and participants of the 2018 Conference on Identification in Econometrics for useful comments. This paper is dedicated to the memory of Gary Chamberlain, whose research motivated this investigation. Address correspondence to Juan Carlos Escanciano, Department of Economics, Universidad Carlos III de Madrid, Getafe, Spain; e-mail: jescanci@eco.uc3m.es.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

2010). This paper aims to establish general conditions for regular and irregular *semiparametric* identification (or lack thereof) and to relate these conditions to the concept of statistical information or the generalizations proposed herein. The results obtained for irregular identification have important practical implications, as any inferences on such parameters are expected to be unstable in empirical analysis. In particular, if a parameter is irregularly identified, then no regular estimator with a parametric rate of convergence exists (see Chamberlain, 1986).[1]

An important observation for relating identification and information is that identification depends on both linear and nonlinear effects (see Sargan, 1983; Chen et al., 2014), while statistical information pertains only to linear effects. To establish a useful link between the two, some structure is thus necessary. Moreover, I show that it is hard to give sufficient conditions for identification in nonlinear models allowing for plausibly high levels of irregularity (see Section 4.1). These arguments motivate an initial focus on linear models, i.e., models with densities that are linear in an infinite-dimensional parameter. In these models, a more complete and transparent analysis of semiparametric identification is possible, permitting both nonparametric unidentification and high degrees of irregularity. Specifically, I establish necessary and sufficient conditions for regular and irregular semiparametric identification. Because many important economic models are not linear but can be written as linear after reparametrization, or by fixing some parameters, the results obtained for linear models are widely applicable.

The analysis of linear models already makes explicit that the separation of irregular identification from no identification is a rather delicate issue. Classical Fisher information is not useful when identification is not regular, because it cannot distinguish between irregular identification and no identification (it is zero in both cases). This paper introduces a new "generalized Fisher information" that seems well-suited for irregular cases. If positive, it implies semiparametric irregular identification when the classical Fisher information is zero and other conditions hold. If zero, it implies impossibility results on rates of convergence for estimators, extending Chamberlain's (1986) impossibility result to slower rates of convergence than parametric.

The identification results are generalized to semiparametric models that are nonlinear in the parameter of interest but linear in nuisance parameters. Examples include commonly used linear and nonlinear panel data models, random coefficient models such as the Mixed Logit and structural models of unemployment duration (see, e.g., Heckman and Singer, 1984a, b), among many others. In this setting, it is possible to present simple sufficient conditions for identification of the main parameter based on the generalized Fisher information, allowing for moderate

[1] Notable examples of irregularly identified parameters in econometrics include densities, regression functions, and their derivatives evaluated at fixed points of continuous variables; regression discontinuity parameters (see Cattaneo and Escanciano, 2017); binary choice coefficients under Manski's (1975) conditions (see Chamberlain, 1986, 2010); sample selection models (see Heckman, 1990; Andrews and Schafgans, 1998; Goh, 2017); mixed proportional models (see Hahn, 1994; Ridder and Woutersen, 2003); average treatment effects (see Khan and Tamer, 2010); or interaction parameters in triangular systems (see Khan and Nekipelov (2018)).

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

irregularity of the main parameter and arbitrary irregularity for nuisance parameters, even when the full model is nonparametrically unidentified.

As a general rule, the impossibility results derived in this paper on regular identification and rates of convergence hold for general linear and nonlinear models. The sufficient conditions for identification require more structure, though, because nonlinearities can overwhelm linear effects (cf. Chen et al., 2014). Section 1.3 in the online Supplemental Appendix presents sufficient conditions for semiparametric identification in nonlinear models, noting that the problem becomes particularly challenging when the model is nonparametrically unidentified.

To illustrate some key concepts and ideas, I consider the canonical example of the average density functional. This functional appears in the asymptotics of the integrated mean squared error of kernel density estimates, and it has been studied by many authors, including Schweder (1975), Hasminskii and Ibragimov (1978), Pfanzagl (1982), Prakasa Rao (1983), Donoho and Liu (1987), Hall and Marron (1987), Bickel and Ritov (1988), Ritov and Bickel (1990), or more recently in Giné and Nickl (2008), and Cattaneo and Jansson (2019), among others. The goal here is not to obtain new results, but rather using this canonical example to illustrate some of the key ideas. One such idea is that identification and estimation are conceptually different concepts, and while positive (generalized) information is necessary for obtaining certain rates of convergence in estimation, it is not sufficient (cf. Ritov and Bickel, 1990). Achieving certain rates of convergence often requires additional assumptions to those guaranteeing a positive generalized information, an issue that I do not address in this paper.

I demonstrate the usefulness of the theory by deriving new identification results for the median willingness to pay (WTP) in contingent valuation studies. The median WTP is identified under weak support conditions, as shown below, and it is an important parameter in this literature. See Carson and Hanemann (2005) for a detailed survey and Lewbel, Linton, and McFadden (2011) and the references therein for semiparametric estimation of moments of WTP. Using the semiparametric identification strategy of Lewbel (1997), Khan and Tamer (2010) have provided sufficient conditions for the mean WTP to be irregularly identified. I revisit the identification of moments of WTP, and then establish irregular identification of the *median* WTP, which is a new result. The approach followed in the present paper is very different from that taken in Khan and Tamer (2010); see Section 6.1 for further discussion.

A third example is a nonlinear model for a consumption-based asset pricing Euler equation with a nonparametric measurement error in consumption. This example demonstrates how the results of this paper can be applied to conditional moment models. Nonparametric and semiparametric treatments of consumption-based asset pricing models, including Newey and Powell (1988), Chen and Ludvigson (2009), Escanciano and Hoderlein (2010), Lewbel et al. (2011), Chen et al. (2014) and Escanciano et al. (2015), do not deal with measurement error in consumption. Yet accounting for measurement error is vital for empirical studies that use household-level data, as shown in, e.g., Shapiro (1984),

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Altonji and Siow (1987), Runkle (1991), and Alan, Attanasio, and Browning (2009). I obtain new primitive conditions for regular identification of the discount factor and measures of risk aversion under more general specifications of the marginal utility and the measurement error mechanism than previously considered. In particular, I show that identification of the discount factor is more robust to assumptions about the measurement error than is identification of risk aversion measures.

In summary, this paper provides general semiparametric identification and impossibility results for regular identification and irregular rates of estimation, and it shows their utility in some example applications. Inference on irregular parameters is challenging. In particular, establishing rates of convergence can be a cumbersome task, but fortunately, the literature has proposed rate-adaptive estimation and inference methods (see, e.g., Newey, 1997; Andrews and Schafgans, 1998; Khan and Tamer, 2010; Chen and Liao, 2014; Chen and Pouzo, 2015). Rate-adaptive methods are recommended for inference about irregularly identified parameters.

The question of whether irregular identification holds or not, and if so, to what degree, is still of first-order importance, because with irregular identification, all estimation methods, including rate-adaptive methods, are expected to be sensitive to the (unknown) data generating process. This paper demonstrates that the strength of this sensitivity, and thus the quality of identification, can be measured by the Fisher information or the generalizations proposed herein.

The rest of the paper is organized as follows. After a literature review, Section 3 sets the statistical framework and introduces some examples that will be used throughout the paper. Section 4 characterizes identification in linear models and defines the generalized Fisher information. Section 5 analyzes semiparametric models. Section 6 studies two examples in detail: the median WTP and the consumption-based Euler equation with measurement error. Section 7 concludes. The Appendix contains proofs of the main results, and the online Supplemental Appendix includes further discussion on identification conditions for linear and nonlinear models.

2. LITERATURE REVIEW

The identification problem has a long history in economics; see the seminal studies by Koopmans (1949), Hurwicz (1950), Koopmans and Reirsol (1950), Fisher (1966) and Rothenberg (1971). Bekker and Wansbeek (2001) and Dufour and Liang (2014) provide more recent contributions as well as a survey of existing results in parametric settings. Chamberlain (1986) shows that a positive semiparametric Fisher information is necessary for regular estimation in semiparametric models. Donoho and Liu (1987) and Pötscher (2002) provide lower bounds for rates of estimation and risk, respectively, based on the continuity properties (or lack thereof) of a functional of an identified object, like a density. Under the explicit assumption of nonparametric identification, van der Vaart

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

(1991) shows the equivalence between positive semiparametric information and a differentiability condition that is necessary for regular estimation. He briefly discusses an "intuitive" local identification condition, but he does not recognize that this condition may be neither sufficient for identification, as shown in Chen et al. (2014), nor necessary, as shown in Sargan (1983). Bickel et al. (1998, Chapter 6) and Ishwaran (1999) present impossibility results on regularity in some exponential and uniform mixture models. Newey (1990) provides further impossibility theorems.

There are, of course, many papers reporting sufficient conditions for identification in specific nonparametric models; see the comprehensive reviews in Matzkin (2007, 2013) and Lewbel (2019). The closest to this paper is Chen et al. (2014), who provide sufficient conditions for nonparametric local identification and for *regular* semiparametric identification for conditional moment models. These authors recognize the difficulty of studying semiparametric irregular identification (see Chen et al., 2014, p. 796) and do not analyze that case, which is the focus of this paper. The results of the present study also help in interpreting their nonparametric identification conditions in terms of statistical information (see Section 4.1). Chen and Santos (2018) investigate local nonparametric regular overidentification. Khan and Tamer (2010) show irregular identification in two important examples and investigate rate-adaptive inference. Chen and Liao (2014) and Chen and Pouzo (2015) provide general inference results for irregular functionals. Independently of this paper, Bonhomme (2011) studies regular and irregular identification and estimation of average marginal effects in nonlinear panel data models with fixed effects. Also related is the identification analysis of Severini and Tripathi (2006, 2012) in nonparametric IV. These authors characterize the set of linear continuous functionals that are regularly and irregularly identified when the nonparametric structural regression is not necessarily identified. Santos (2011) and Escanciano and Li (2021) investigate semiparametric regular estimation in the IV setting.

The present paper deals broadly with semiparametric identification in likelihood and conditional moment models. Furthermore, it follows the tradition of the seminal work by Rothenberg (1971) in linking the identification problem to the concept of statistical information (and generalizations proposed herein), although in a nonparametric setting.

3. SETTING AND EXAMPLES

The data are an independent and identically distributed sample $Z_1, ..., Z_n$ from a distribution $\mathbb{P}$ that belongs to a class of probability measures $\mathcal{P} = \{\mathbb{P}_\lambda : \lambda \in \Lambda\}$, where Λ is a subset of a Hilbert space $(\mathbf{H}, \langle \cdot, \cdot \rangle_{\mathbf{H}})$, with inner product $\langle \cdot, \cdot \rangle_{\mathbf{H}}$ and norm $\|\cdot\|_{\mathbf{H}}$. For example, in parametric models, $\Lambda \subset \mathbb{R}^m$ and $\|\cdot\|_{\mathbf{H}} = |\cdot|$ is the euclidean norm. This paper focusses on nonparametric models where Λ is infinite-dimensional, e.g., a subset of a space of probability densities. The parameter that generates the data is denoted by $\lambda_0 \in \Lambda$, i.e., $\mathbb{P} = \mathbb{P}_{\lambda_0}$. The goal is to find sufficient and necessary conditions for identification of $\phi(\lambda_0)$, for a functional $\phi : \Lambda \mapsto \mathbb{R}^p$,

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

allowing for the full model $\mathcal{P}$ to be unidentified at λ_0. That is, the equation $\mathbb{P}_\lambda = \mathbb{P}$ may have more than one solution in Λ. This setting includes as a special case semiparametric models where $\lambda = (\theta, \eta) \in \Lambda = \Theta \times H, \Theta \subset \mathbb{R}^p$ and H is a subset of another Hilbert Space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$. A leading example of such a functional is the finite-dimensional parameter, i.e., $\phi(\theta, \eta) = \theta$. However, the setting also includes functionals of the nuisance parameter $\phi(\lambda) = \chi(\eta)$, where $\chi : H \mapsto \mathbb{R}^p$, which, despite the name, may be of interest. For example, η can measure unobserved heterogeneity, and one might be interested in average marginal effects or policy counterfactuals that involve averaging across a heterogeneous population.

To introduce the definition of identification, let f_λ be the density of $\mathbb{P}_\lambda$ with respect to (wrt) a σ-finite measure μ. Denote by $\mathcal{B}_\delta(\lambda_0) = \{\lambda \in \Lambda : \|\lambda - \lambda_0\|_{\mathbf{H}} < \delta\}$ a ball of radius δ around λ_0.

DEFINITION (Semiparametric identification). $\phi(\lambda)$ is locally identified in $\mathcal{P}$ at $\phi(\lambda_0)$ if there exists $\delta > 0$ such that, for all $\lambda \in \mathcal{B}_\delta(\lambda_0)$, $f_\lambda = f_{\lambda_0} \mu$-almost surely ($\mu$-a.s.) implies $\phi(\lambda) = \phi(\lambda_0)$. If this implication holds for all $\lambda \in \Lambda$, then $\phi(\lambda)$ is (globally) identified at $\phi(\lambda_0)$.

To simplify the exposition, I simply write "$\phi(\lambda_0)$ is locally identified" rather than "$\phi(\lambda)$ is locally identified in $\mathcal{P}$ at $\phi(\lambda_0)$," and if "locally" is dropped, then identification is meant to be global. For parametric models, i.e., $\Lambda \subset \mathbb{R}^m$, Fisher (1966) and Rothenberg (1971) show that, with sufficient smoothness of the model, nonsingularity of the Fisher information matrix is necessary and sufficient for local identification of λ_0. In nonparametric models, positive information is not necessary for identification anymore, and this leads to the classification of identification as regular and irregular (cf. Khan and Tamer, 2010).

DEFINITION (Regular and irregular semiparametric identification). $\phi(\lambda_0)$ is (locally) regularly (respectively, irregularly) identified if it is (locally) identified, and its Fisher information is positive (respectively, zero).

Identification and regularity/irregularity are separate concepts. So, for example, the negation of regular identification, which is used extensively throughout the paper, entails two possibilities: irregular identification or no identification at all.

The main ideas and results will be illustrated with several examples.

Example 1 (Average density). A canonical functional is the average density

$$\phi(\lambda_0) = \int_{\mathbb{R}} \lambda_0^2(z) dz \equiv \mathbb{E}_{\lambda_0}[\lambda_0(Z_i)],$$

where λ_0 is the (Lebesgue) density of Z_i, and where $\mathbb{E}_{\lambda_0}$ denotes expectation under the density λ_0. An even simpler functional is the average functional $\phi(\lambda_0) = \mathbb{E}_{\lambda_0}[w(Z_i)]$ for a known measurable function w. The nonparametric model is simply $\mathcal{P} = \{f_\lambda = \lambda : \lambda \in \Lambda\}$, for a subset Λ of densities contained in $\mathbf{H}$, where $\mathbf{H}$ is, for example, a Hilbert space of square integrable measurable functions on the real line. Identification is obviously of no concern here, as λ_0 is identified, but this canonical example is still useful to illustrate some of the key ideas of the paper. ▲

Example 2 (Willingness to pay). In contingent valuation studies, one observes $Z_i = (Y_i, V_i, X_i)$, where $Y_i = 1(W_i \leq V_i)$, i.e., $Y_i = 1$ if $W_i \leq V_i$, and zero otherwise,

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

V_i is a continuous random variable chosen by the researcher, with known distribution F_V, and X_i a d-dimensional vector of covariates. Here, W_i is WTP of individual i for a new product or resource, which is an unobserved continuous nonnegative random variable. It is assumed that W_i and V_i are conditionally independent given X_i. The density of Z_i, wrt a suitable measure μ, is

$$f_{\lambda_0}(y,v,x) = [G_0(v,x)]^y [1 - G_0(v,x)]^{1-y},$$

where $G_0(v,x) = \mathbb{P}[W_i \leq v | X_i = x]$ and $\lambda_0(v,x) = \partial G_0(v,x)/\partial v$. Here, one parameter of interest is the median of the unconditional distribution of W_i,

$$\phi(\lambda_0) = Median(W_i).$$

Lewbel (1997) and Lewbel et al. (2011) investigate nonparametric and semiparametric estimation of moments $\phi(\lambda_0) = \mathbb{E}[r(W_i)]$ and $\phi(\lambda_0) = \mathbb{E}[r(W_i, X_i)]$, respectively. Khan and Tamer (2010) show that $\mathbb{E}[W_i]$ is irregularly identified when the support of W_i is unbounded, and discuss rates of convergence for this functional. There is also an extensive literature for the related binary choice model when W_i has the representation $W_i = \theta_0' X_i + \varepsilon_i$. In Section 6.1 of this paper, I study identification of the median WTP. This example is also useful to illustrate the systematic aspect of the proposed method, i.e., a single approach can be used for different functionals. ▲

In the following example, I obtain new regular identification results by an application of the characterization of regular semiparametric identification to a conditional moment model.

Example 3 (Consumption-based asset pricing models with measurement error). Consumption-based asset pricing Euler equations are important models in economics. When applied to microeconomic data, it is vital to account for measurement error in consumption, as in

$$\mathbb{E}\left[\theta_0 \dot{u}_0(C_{t+1}^*) R_{t+1} - \dot{u}_0(C_t^*) \middle| \mathcal{F}_t\right] = 0,$$

where θ_0 is the discount factor, $\dot{u}_0$ is the marginal utility of consumption C_t^*, R_{t+1} is the gross return of an asset, and $\mathcal{F}_t$ denotes the σ-field generated by the agent's information set at time t. The econometrician observes C_t, which is a noisy measure of C_t^*, following the specification

$$C_t = m(C_t^*, \varepsilon_t),$$

where m is unknown and ε_t is the measurement error. The primitives of the model, λ_0, are $\theta_0, \dot{u}_0, m$, and the distribution of $(C_{t+1}^*, R_{t+1}, \varepsilon_{t+1}, C_t^*, \varepsilon_t)$ given $\mathcal{F}_t$. Of particular interest are θ_0 and $\dot{u}_0$. The observed data are $Z_i = (C_{t+1,i}, C_{t,i}, R_{t+1,i}, X_{t,i})$, for a sample of households, and where $X_{t,i}$ is a vector of household characteristics (e.g., family size) in $\mathcal{F}_{t,i}$. The results of this paper are applied to this example to obtain sufficient conditions for identification of the discount factor, $\phi(\lambda_0) = \theta_0$, and the Average Arrow–Pratt coefficient of Absolute Risk Aversion (AARA), $\phi(\lambda_0) = \mathbb{E}\left[\left(-\partial \dot{u}_0(C_t^*)/\partial C_t^*\right)/\dot{u}_0(C_t^*)\right]$. This example is studied in detail in Section 6.2. ▲

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

4. LINEAR NONPARAMETRIC MODELS

This section first introduces some notation that will be used throughout the paper. For a generic measure ν, let $L_q(\nu), q \geq 1$, denote the Banach space of (equivalence classes of) real-valued measurable functions h such that $\|h\|_{q,\nu} := \left(\int |h|^q d\nu\right)^{1/q} < \infty$ (henceforth, I drop the sets of integration in integrals and the qualification ν-almost surely for simplicity of notation). So, for example, a function in $L_q(\nu)$ is discontinuous when there is no continuous function in its equivalence class. Define the Hilbert space L_2 of $\mathbb{P}$-square integrable measurable functions with inner product $\langle h,f\rangle = \int hfd\mathbb{P}$ and norm $\|h\|^2 = \langle h,h\rangle$ (I drop the dependence on $q=2$ and $\nu = \mathbb{P}$ in this case). The set $L_q^0(\nu)$ (L_q^0) is the subspace of zero-mean functions in $L_q(\nu)$ (respectively, L_q). Henceforth, for a generic linear operator $K : \mathcal{G}_1 \to \mathcal{G}_2$, $\mathcal{N}(K) := \{f \in \mathcal{G}_1 : Kf = 0\}$ denotes its kernel.

To derive the results, it is convenient to give a linear structure to the set of deviations $\lambda - \lambda_0$ from λ_0. Define $B_0 = \{b \in \mathbf{H} : \lambda_0 + b \in \Lambda\}$and let $T(\lambda_0)$ denote the linear span of elements in B_0. The identification results given below will be relative to the tangent set $T(\lambda_0)$ (i.e., relative to Λ). See Remark 1 below for a convenient reparametrization when Λ is a subset of densities.

The goal is to relate identification with the concept of statistical information in a nonparametric setting. To that end, let us consider the *score operator* (see, e.g., Begun et al., 1983), which, for linear models defined as in Assumption 1 below, is the operator $S : T(\lambda_0) \mapsto L_2$ given by

$$Sb \equiv S_{\lambda_0} b := \frac{f_{\lambda_0+b} - f_{\lambda_0}}{f_{\lambda_0}} 1(f_{\lambda_0} > 0). \tag{1}$$

Note that f_{λ_0+b} is well defined, when $b \in B_0$, since $\lambda_0 + b \in \Lambda$, and I provide conditions under which f_{λ_0+b} is also well defined, for $b \notin B_0$. Specifically, under linearity of S, this operator has a unique extension from B_0 to $T(\lambda_0)$ (see Debnath and Mikusinski, 2005, p. 26), and hence Sb is well defined when $b \in T(\lambda_0) \setminus B_0$. The condition $Sb \in L_2$ may impose additional assumptions on the model, as the average density example illustrates.

Example 1 (Cont.). In this example, since $f_\lambda = \lambda$, the score operator is simply

$$Sb = \frac{b}{\lambda_0} 1(\lambda_0 > 0).$$

The condition $Sb \in L_2$ then requires

$$\int_{\lambda_0(z)>0} \frac{b^2(z)}{\lambda_0(z)} dz < \infty.$$

This is similar to the assumption of finite Fisher information for location (see Bickel et al., 1998). ▲

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

More generally, existence of the score operator is necessary for the classical mean squared differentiability assumption, which requires that for every path $\lambda_t \in \Lambda$ with $t^{-1}(\lambda_t - \lambda_0) \to b \in T(\lambda_0) \subset \mathbf{H}$, the following holds:

$$\left\| \frac{f_{\lambda_t}^{1/2} - f_{\lambda_0}^{1/2}}{t} - \frac{1}{2} S b f_{\lambda_0}^{1/2} \right\|_{2,\mu} \to 0 \text{ as } t \downarrow 0. \tag{2}$$

The definition of Sb in (2) is the most commonly used in the literature and applies equally to linear and nonlinear models. I will use this latter definition for nonlinear models, but keep the simpler and more natural definition in (1) for linear models. Often, the path can be taken of the form $\lambda_t = \lambda_0 + tb$, and $Sb = \partial \log f_{\lambda_0 + tb}/\partial t$ is simply the score associated with the parametric submodel $f_{\lambda_0 + tb}$ at the "truth" $t = 0$ (corresponding to λ_0) in the "direction" b, where, henceforth, derivatives wrt to t are one-sided and evaluated at zero.

This section investigates identification when the density f_λ and the functional ϕ are linear.[2] To simplify the exposition, it is assumed that the functional is a scalar, with the understanding that all the results below have straightforward extensions to multivariate functionals.

Assumption 1. (i) The map $\dot{\phi} : T(\lambda_0) \subseteq \mathbf{H} \mapsto \mathbb{R}$ defined by $\dot{\phi}(b) = \phi(\lambda_0 + b) - \phi(\lambda_0)$ is linear; (ii) the score operator $S : T(\lambda_0) \subseteq \mathbf{H} \mapsto L_2$ in (1) is well defined and linear; (iii) for each $b \in \mathcal{N}(S)$, there exists $c \equiv c(b) \in \mathbb{R}$, $c \neq 0$, such that $\lambda_0 + cb \in \Lambda$; (iv) $\dot{\phi}$ and S are continuous.

Assumption 1(i) holds for the leading example of the finite-dimensional parameter in a semiparametric model, i.e., $\phi(\theta, \eta) = \theta$. It also holds for the average functional of Example 1 with a known w, but it does not hold when w depends on λ_0, as with the average density. Nevertheless, for identified smooth nonlinear functionals, regularity or irregularity depends only on the linear approximation, so some of our results based on Assumption 1 are still applicable to these cases. Section 1.3 in the online Supplemental Appendix relaxes Assumption 1(i). Assumption 1(ii) holds in Examples 1 and 2. In other models, the linearity assumption holds after a suitable reparametrization. For example, consider a simple model of a binary outcome Y_i^* that is only observed when $D_i = 1$. That is, the available data are $Z_i = (Y_i, D_i, X_i)$, where $Y_i = Y_i^* D_i$ and Y_i^* is independent of D_i given X_i. Define $q(X_i) = \mathbb{E}\left[Y_i^* \middle| X_i \right]$ and $p(X_i) = \mathbb{E}[D_i | X]$. The density of (Y_i, D_i, X_i) is a nonlinear function of p and q, since, for example, $\mathbb{P}(Y_i = 1, D_i = 1 | X_i = x) = q(x)p(x)$, but if we reparametrize the density in terms of $\lambda_0 = (\lambda_{01}, \lambda_{02})$ with $\lambda_{01}(x) = q(x)p(x)$ and $\lambda_{02}(x) = p(x)$, then the density becomes linear in λ_0. Section 5 below and Section 1.3 in the online Supplemental Appendix relax Assumption 1(ii). Assumption 1(iii) is only used to prove the necessity of the main identification condition below, and it can be dropped altogether by restricting attention only to $b \in B_0$, although it facilitates exposition. Overall, Assumption 1 is convenient, because with this assumption, identification can be fully characterized. Thus, the identification results under

[2] For the sake of exposition, I refer to these as linear, although a more mathematically precise name is affine.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Assumption 1 provide a benchmark for what can be achieved in more complicated situations.

Assumption 1(iv) guarantees the existence of the adjoint operator $S^*: L_2 \mapsto T(\lambda_0)$ of S, satisfying $\langle g, Sb\rangle = \langle S^*g, b\rangle_{\mathbf{H}}$, for all $g \in L_2$ and $b \in T(\lambda_0)$. The following definition extends the Fisher information matrix to a nonparametric context (cf. Koševnik and Levit, 1976).

DEFINITION (Fisher information). The information operator is defined as $I_{\lambda_0} := S^*S$.

Roughly, $I_{\lambda_0}b$ measures the Fisher information of λ_0 in the direction $b \in T(\lambda_0)$, i.e., the classical Fisher information corresponding to the "parametric" model f_{λ_0+tb} at $t = 0$ in the direction b. To establish a link between I_{λ_0} and identification, note that, under Assumption 1, semiparametric identification of $\phi(\lambda_0)$ will hold if, for all $b \in T(\lambda_0)$,

$$f_{\lambda_0+b} - f_{\lambda_0} \equiv f_{\lambda_0} 1(f_{\lambda_0} > 0) Sb = 0 \Longrightarrow \phi(\lambda_0 + b) - \phi(\lambda_0) \equiv \dot{\phi}(b) = 0, \tag{3}$$

or, since $f_{\lambda_0} > 0$ $\mathbb{P}$-a.s. and $\mathcal{N}(S) = \mathcal{N}(I_{\lambda_0})$,

$$\mathcal{N}(I_{\lambda_0}) \subset \mathcal{N}(\dot{\phi}). \tag{4}$$

From the left side of (3), it is clear why $b \in \mathcal{N}(S)$ are problematic "directions"—they correspond to observationally equivalent parameters $\lambda_0 + b$ and λ_0. Thus, (4) requires that observationally equivalent parameters lead to the same functional. The following proposition proves that (4), which involves the nonparametric Fisher information, is necessary and sufficient for semiparametric identification of $\phi(\lambda_0)$ under Assumption 1.

PROPOSITION 4.1. *Under Assumption 1, identification of $\phi(\lambda_0)$ holds iff (4) holds.*

Without Assumption 1(ii), both implications of Proposition 4.1 fail, which motivates the initial focus on linear models. That (4) is not sufficient for identification follows from a counterexample I give in the online Supplemental Appendix, which builds on one given in Chen et al. (2014), whereas the fact that it is not necessary follows from Sargan (1983). Heuristically, identification depends, in general, on both linear and nonlinear effects, while (4) is about linear effects only. Some structure is thus needed for the intuitive identifiability condition (4) to be useful for identification. Assumption 1 is a natural starting point, because with this assumption, identification is characterized.

The next remark points out that the results above can be extended to more general settings without a Hilbert space structure. Furthermore, when the parameter λ is a density, as it happens in many applications, a more convenient definition of derivatives is useful.

Remark 1. (i) When Λ is a subset of densities wrt a σ-finite measure π, it is natural to define the score operator as an operator from $L_1(\pi)$ to $L_1(\mathbb{P})$, since $\lambda \in L_1(\pi)$ and $Sb \in L_1(\mathbb{P})$ without additional assumptions. Proposition 4.1 holds

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

with these extended definitions as well. (ii) If a Hilbert space approach is preferred when Λ is a subset of densities, it is convenient to introduce directions $b \in L_2^0(G_0)$, where G_0 is the measure associated with λ_0, and redefine the score operator in (1) and the functional $\dot{\phi}$ with b replaced by $\lambda_0 b$ (and $T(\lambda_0)$ as the linear span of $B_0 = \{b \in L_2^0(G_0) : \lambda_0 + \lambda_0 b \in \Lambda\}$). The specific definition used for the operator S and functional $\dot{\phi}$ will be clear from the context. To provide a useful interpretation, note $b = \partial \log \lambda_t / \partial t$ for $\lambda_t = \lambda_0(1 + tb)$. Thus, with this parametrization, the score operator transforms scores of Λ into scores of the model $\mathcal{P}$.

Example 1 (Cont.). Applying the reparametrization of Remark 1(ii) to the average density example, we obtain the score operator $Sb = b1(\lambda_0 > 0)$. The condition $Sb \in L_2$ is then naturally satisfied (since $b \in L_2^0$). Furthermore, since $\mathcal{N}(S) = \{0\}$, any linear continuous functional is identified (as expected). ▲

The identification condition (4) is based on the nonparametric Fisher information, and, as such, it is not useful if the goal is to disentangle regular and irregular semiparametric identification. To introduce a more useful characterization, I first define the semiparametric Fisher information for ϕ. The information for estimating the parameter $\psi(t) = \phi(\lambda_t), \lambda_t := \lambda_0 + tb$, under the density f_{λ_t} at $t = 0$ is, by the delta method, equal to $[\partial\psi(t)/\partial t]^{-1} ||Sb||^2 [\partial\psi(t)/\partial t]^{-1} = ||Sb||^2/[\dot{\phi}(b)]^2$. The semiparametric Fisher information is the infimum of the information over all such parametric submodels (cf. Stein, 1956) and is given by

$$I_\phi = \inf_{b \in \mathcal{B}_\phi} \frac{||Sb||^2}{\left[\dot{\phi}(b)\right]^2}, \tag{5}$$

where $\mathcal{B}_\phi := \left\{b \in T(\lambda_0) : \dot{\phi}(b) \neq 0, \left|\dot{\phi}(b)\right| \leq 1\right\}$.

By the continuity in Assumption 1(iv), S and $\dot{\phi}$ are uniquely extended to $\overline{T(\lambda_0)}$, where henceforth, for a subspace V, $\overline{V}$ denotes the closure of V in the norm topology. Moreover, there exists an $r_\phi \in \overline{T(\lambda_0)}$, called the Riesz representer of $\dot{\phi}$, such that, for all $b \in \overline{T(\lambda_0)}$,

$$\dot{\phi}(b) = \langle b, r_\phi \rangle_{\mathbf{H}}.$$

I can then identify $\dot{\phi}$ with r_ϕ, and provide identification results in terms of r_ϕ using duality. I will make clear in the text when S is viewed as acting on $T(\lambda_0)$ or on its extension $\overline{T(\lambda_0)}$. Henceforth, whenever identification is discussed in terms of r_ϕ, as in the next theorem, I mean S and $\dot{\phi}$ to be defined on $\overline{T(\lambda_0)}$ (so r_ϕ is guaranteed to exist by the Riesz representation theorem).

I illustrate the important concept of Riesz representer in the context of the nonparametric density model of Example 1.

Example 1 (Cont.). In the setting of Example 1, consider the average functional $\phi(\lambda_0) = \mathbb{E}_{\lambda_0}[w(Z_i)]$ for a known function w satisfying $\mathbb{E}_{\lambda_0}\left[w^2(Z_i)\right] < \infty$. Consider the nonparametric model where $\overline{T(\lambda_0)} = L_2^0$. Then, it is straightforward to show

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

$\dot{\phi}(b) = \phi(\lambda_0 + \lambda_0 b) - \phi(\lambda_0) = \langle b, w \rangle$, so for $\mathbf{H} = L_2$, it holds that $r_\phi = w - \mathbb{E}_{\lambda_0}[w(Z_i)]$. Furthermore, by Cauchy–Schwarz

$$I_\phi = \inf_{b \in \mathcal{B}_\phi} \frac{||b||^2}{[\langle b, w \rangle]^2} \geq \frac{1}{||r_\phi||^2},$$

and by taking $b = r_\phi / ||r_\phi||$, equality is achieved in the last display, so $I_\phi = 1/var(w(Z_i))$. This is a well-known result in the literature. The average density functional $\phi(\lambda_0) = \mathbb{E}_{\lambda_0}[\lambda_0(Z_i)]$ is nonlinear, with $\phi(\lambda_0 + \lambda_0 b) - \phi(\lambda_0) = \langle b, 2\lambda_0 \rangle + \langle b^2, \lambda_0 \rangle$. Applying the definition of Fisher information and Riesz representer to the linear term $\dot{\phi}(b) = \langle b, 2\lambda_0 \rangle$, as usual, I obtain the well-known efficiency bound $var(2\lambda_0(Z_i))$ and influence function $r_\phi = 2\lambda_0 - \mathbb{E}_{\lambda_0}[2\lambda_0(Z_i)]$).▲

Let $\mathcal{R}(S^*) := \{f \in \overline{T(\lambda_0)} : \exists g \in L_2, S^* g = f\}$. Then, the following result provides a full characterization of semiparametric regular and irregular identification in linear models.

THEOREM 4.1. *Under Assumption 1: (i) $\phi(\lambda_0)$ is regularly identified iff $r_\phi \in \mathcal{R}(S^*)$; (ii) $\phi(\lambda_0)$ is irregularly identified iff $r_\phi \in \overline{\mathcal{R}(S^*)} \setminus \mathcal{R}(S^*)$; and (iii) $\phi(\lambda_0)$ is unidentified iff $r_\phi \notin \overline{\mathcal{R}(S^*)}$.*

The results that seem to be novel here are the "identification" parts and the separation of irregular identification from no identification. The fact that $r_\phi \in \mathcal{R}(S^*)$ is equivalent to $I_\phi > 0$ is due to van der Vaart (1991, Theorem 4.1). As discussed earlier, the characterizations of Theorem 4.1 do not necessarily hold without Assumption 1, and some care must be exercised to extrapolate these results to nonlinear models. Section 1.3.1 in the online Supplemental Appendix presents a nonlinear counterexample, building on that given in Chen et al. (2014), that shows that $I_\phi > 0$ may hold while $\phi(\lambda_0)$ is not identified. Severini and Tripathi (2006, 2012) obtain related results to Theorem 4.1 for the specific problem of linear functionals in the nonparametric IV setting.

I illustrate some key ideas with the average density example.

Example 1 (Cont.). In the average density example with a nonparametric tangent space, the score operator $S : L_2^0 \mapsto L_2$ is $Sb = b$, and hence, $S^* b(z) = b(z) - \mathbb{E}_{\lambda_0}[b(Z_i)]$. Then, clearly $r_\phi = 2\lambda_0 - \mathbb{E}_{\lambda_0}[2\lambda_0(Z_i)])$ belongs to $\mathcal{R}(S^*)$ if $\lambda_0 \in L_2$. That is, identification of the average density is regular if $var(\lambda_0(Z_i)) < \infty$. In important work, Ritov and Bickel (1990) have shown that although the Fisher information is positive, regular $\sqrt{n}$-estimation requires stronger conditions than the minimal conditions for positive Fisher information. However, it is important to note that regular identification and regular estimation are different concepts. While a positive Fisher information is necessary for regular estimation, it may not be sufficient. In this paper, I do not study achievable rates of convergence, which is beyond the scope of this paper and has been well studied in the literature for this example (see, e.g., Bickel and Ritov, 1988; Giné and Nickl, 2008). Irregular identification of the average density holds when $var(\lambda_0) = \infty$. Thus, with the

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

condition that the functional is well defined, $\lambda_0 \in L_q$, for some $q, 1 \leq q < 2$, and $\lambda_0 \notin L_2$ correspond to irregular identification of the average density.▲

4.1. Irregular Identification and Generalized Information

When the Fisher information I_ϕ is zero, it does not provide information on identification (it cannot distinguish between irregular identification and no identification). This section introduces a "generalized Fisher information" that extends, in a sense described later, the classical Fisher information to irregular cases, and which is given by

$$I_{\phi,\rho} = \inf_{b \in \mathcal{B}_\phi} \frac{||Sb||^2}{\left|\dot{\phi}(b)\right|^{2\rho}}, \tag{6}$$

where $1 \leq \rho < \infty$. The classical Fisher information corresponds to $\rho = 1$, i.e., $I_\phi \equiv I_{\phi,1}$. Furthermore, it can be shown that $I_{\phi,1} \leq I_{\phi,\rho}$, for $1 < \rho < \infty$.

The sense in which the generalized Fisher information provides a generalization of the classical Fisher information is shown in the next two results. The first result extends the sufficient condition for identification in Theorem 4.1(i) to the irregular case. Under Assumption 1, $I_{\phi,1} = 0$ and $I_{\phi,\rho} > 0$, for some $\rho > 1$, corresponds to irregular identification. Because $\left|\dot{\phi}(b)\right| \leq I_{\phi,\rho}^{-2\rho} ||Sb||^{1/\rho}$, for $b \in \mathcal{B}_\phi$, an interpretation of a positive generalized information is that $\dot{\phi}(b)$ is continuous in $\mathcal{B}_\phi$ wrt the Fisher semi-norm $||Sb||$, with a modulus of continuity quantified by $1/\rho$ (smaller ρ corresponding to more regularity). The inequality above directly gives identification on a restricted set, and the next result proves that this set can be extended to the whole parameter space Λ under Assumption 1.

THEOREM 4.2. *Let Assumption 1(i–ii) holds. If $I_{\phi,\rho} > 0$, for some $\rho, 1 \leq \rho < \infty$, then $\phi(\lambda_0)$ is identified.*

As discussed at the end of this section, there is an extensive literature on nonparametric identification that requires, implicitly or explicitly, the condition $I_{\phi,\rho} > 0$ to establish rates of estimation. Theorem 4.2 shows $I_{\phi,\rho} > 0$ is sufficient for identification, whereas the next result shows it is necessary for achieving certain rates of convergence. Let $\mathcal{A}$ denote a class of sequences in Λ, and let $\mathbb{P}_\lambda^n$ denote the n-fold probability of $\mathbb{P}_\lambda$. I provide a formal definition of rate of convergence (see, e.g., Ishwaran, 1996, Definition 7).

DEFINITION (Rate of convergence). The estimator T_n has a rate of convergence r_n on $\mathcal{A}$ for estimating $\phi(\lambda_0)$ if, for each $\varepsilon > 0$, there exists $K(\varepsilon) > 0$ such that, for each $\{\lambda_n\} \in \mathcal{A}$,

$$\limsup_{n \to \infty} \mathbb{P}_{\lambda_n}^n \left(|T_n - \phi(\lambda_n)| > K(\varepsilon) r_n\right) < \varepsilon.$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

THEOREM 4.3. *Suppose that, for each $\varepsilon > 0$, there exists a path $\lambda_t \in \Lambda$ passing through λ_0 such that, for all t, sufficiently small the following holds: (i) $Ct \leq |\phi(\lambda_t) - \phi(\lambda_0)| \leq 1$, for some $C > 0$, and (ii) $\left\| \left(f_{\lambda_t} - f_{\lambda_0}\right)/f_{\lambda_0} \right\|^2 < \varepsilon t^{2\rho}$, for $1 \leq \rho < \infty$. Suppose $\mathcal{A}$ contains all sequences $\{\lambda_n\}$ for which $\phi(\lambda_n) = \phi(\lambda_0) + O(n^{-1/2\rho})$. Then, the rate of convergence for any estimator of $\phi(\lambda_0)$ on $\mathcal{A}$ must be slower than $O(n^{-1/2\rho})$.*

Conditions (i) and (ii) in Theorem 4.3 correspond to $I_{\phi,\rho} = 0$, since under these conditions, $b_t \equiv (\lambda_t - \lambda_0) \in \mathcal{B}_\phi$ and

$$I_{\phi,\rho} = \inf_{b \in \mathcal{B}_\phi} \frac{||Sb||^2}{[\dot{\phi}(b)]^{2\rho}} \leq \frac{\varepsilon}{C^{2\rho}},$$

and because $\varepsilon > 0$ is arbitrary, it must hold that $I_{\phi,\rho} = 0$. To relate (ii) with the classical mean squared differentiability, use the basic inequality $(\sqrt{a} - \sqrt{b})^2 \leq (b-a)^2/a$, for any $a, b \geq 0$, and note that (ii) then implies for t sufficiently small

$$\left\| \frac{f_{\lambda_t}^{1/2} - f_{\lambda_0}^{1/2}}{t} \right\|_{2,\mu}^2 < \varepsilon t^{2(\rho-1)}.$$

Thus, condition (ii) can be viewed as a mean squared differentiability assumption with a zero derivative condition at λ_0 and a Lipschitz bound on the derivative in a neighborhood of λ_0.

I stress that Assumption 1 is not required in Theorem 4.3, so the theorem holds for linear and nonlinear models/parameters. Theorem 4.3 extends the impossibility result of Chamberlain (1986) to the irregular case $\rho > 1$. Its proof relies on techniques, originally due to LeCam (1973), that bound the total variation distance between the distribution under f_{λ_n} and that under f_{λ_0}, for a suitable sequence λ_n. Previously, Donoho and Liu (1987) have used similar techniques based on Hellinger distance to provide lower bounds for convergence of estimators. The modest contribution of Theorem 4.3 is to connect the zero generalized Fisher information introduced in this paper with these lower bounds, extending important work by Chamberlain (1986) to the irregular case.

It is important to stress that there is nothing that prevents the possibility that $I_{\phi,\rho} = 0$, for all $\rho \geq 1$, which, under Theorem 4.3, implies impossibility of polynomial rates. Indeed, in several useful models, logarithmic rates are common; see, e.g., Fan (1991) for classical measurement error problems. This possibility suggests that the definition of the generalized information and the conditions of Theorem 4.3 should be modified to accommodate severely irregular cases. For a function ψ that is increasing, nonnegative, right continuous at 0 and with $\psi(0) \downarrow 0$, one can define the generalized Fisher information

$$I_{\phi,\psi} = \inf_{b \in \mathcal{B}_\phi} \frac{||Sb||^2}{\psi\left([\dot{\phi}(b)]^2\right)}.$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

With this modification, different degrees of irregularity, including severe irregularity, are allowed. The case $\psi(\epsilon) = \epsilon^{\rho}, 1 < \rho \leq 2$, corresponds to mild or moderate irregularity, while $\psi(\epsilon) = \exp(\epsilon) - 1$ or $\psi(\epsilon) = \exp(-1/(\epsilon^{a}))$, with $a > 0$, is suitable for severe irregularity with possibility of logarithmic rates. A version of Theorem 4.3 that allows for severe irregularity follows *mutatis mutandis*, simply replacing $t^{2\rho}$ by $\psi\left(t^{2}\right)$.

At this point, it is useful to compare the results of this paper with the general nonparametric local identification results in Chen et al. (2014) for conditional moment restrictions models. These authors obtain sufficient conditions for nonparametric identification of linear and nonlinear conditional moments by suitably restricting the parameter space. When conditional moments are only Frechet differentiable, they consider the parameter space to have tangents in $\{b : ||\dot{m}b||^2 > C\|b\|_{\mathbf{H}}^2\}$, for the derivative $\dot{m}$ of a conditional mean operator m and a positive constant C. In our setting, an analog that allows comparison with statistical information would be $m(\lambda) = \left(f_{\lambda} - f_{\lambda_0}\right)/f_{\lambda_0}$, with derivative $\dot{m}(\lambda) = S(\lambda - \lambda_0)$.[3] On the parameter space with tangents $\{b : ||Sb||^2 > C\|b\|_{\mathbf{H}}^2\}$, the nonparametric information is positive (i.e., regular nonparametric identification), which implies a positive semiparametric Fisher information for *all* linear continuous functionals. Chen et al. (2014) also consider conditions corresponding to higher-order differentiability, and these conditions do allow for irregular semiparametric identification. In their general case, they restrict tangents to the set $\{b : ||Sb||^2 > C\|b\|_{\mathbf{H}}^{2\rho}\}$, for $\rho > 1$, which implies a positive generalized Fisher information $I_{\phi,\rho}$ for *all* continuous linear functionals ϕ. To prove such a result, use $\left|\dot{\phi}(b)\right| \leq \left\|r_{\phi}\right\|_{\mathbf{H}}\|b\|_{\mathbf{H}}$ and their assumption $\mathcal{B}_{\phi} \subset \{b : ||Sb||^2 > C\|b\|_{\mathbf{H}}^{2\rho}\}$ to bound

$$I_{\phi,\rho} = \inf_{b \in \mathcal{B}_{\phi}} \frac{||Sb||^2}{[\dot{\phi}(b)]^{2\rho}} \geq \frac{C}{\left\|r_{\phi}\right\|_{\mathbf{H}}^{2\rho}} > 0.$$

This shows that the restrictions on neighborhoods in Chen et al. (2014) have a statistical interpretation in terms of the generalized Fisher information introduced here (for all continuous linear functionals). The parameter ρ is also linked to the nonlinearity permitted in the model (see Assumption 2 in Chen et al., 2014 or Assumption 2 below), which typically restricts its values to $1 \leq \rho \leq 2$. The case $\rho > 2$ only holds when the second derivative of $m(\lambda)$ is zero. From this discussion, it seems that it may be hard to accommodate nonlinear cases for severely ill-posed problems where $I_{\phi,\rho} = 0$, for all $\rho \geq 1$ and *some* functional ϕ. Examples of such severe irregularity for certain functionals ϕ include nonparametric IV models or measurement error models with Gaussian errors.

[3] More precisely, the effective score operator in their conditional moment restriction model is $S(\lambda - \lambda_0) = \Sigma^{-1}\dot{m}(\lambda - \lambda_0)$ for a conditional variance Σ. Under the standard assumption that the eigenvalues of Σ are bounded away from zero, there is a one-to-one relation between identification in conditional moment models as in Chen et al. (2014) and identification in terms of statistical information as considered here (see, e.g., Chamberlain, 1992).

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

To give sufficient conditions for irregular identification that allow for a variety of degrees of irregularity for different functionals, consider adding the following condition to Assumption 1.

Assumption 1. (v) The score operator S is compact.

Assumption 1(v) guarantees the existence of a sequence $\{\lambda_j, \varphi_j, \psi_j\}_{j=1}^{\infty}$ such that (cf. Kress, 1999, Theorem 15.16)

$$S\varphi_j = \lambda_j \psi_j \qquad \text{and} \qquad S^*\psi_j = \lambda_j \varphi_j. \tag{7}$$

This is the so-called singular value decomposition (SVD) of S. The elements $\{\varphi_j\}_{j=1}^{\infty}$ and $\{\psi_j\}_{j=1}^{\infty}$ are complete orthonormal bases for $\overline{\mathcal{R}(S^*)}$ and $\overline{\mathcal{R}(S)}$, respectively, and the singular values λ_j are the squared-root eigenvalues of the information operator $I_{\lambda_0} = S^*S : \overline{T(\lambda_0)} \mapsto \overline{T(\lambda_0)}$. Furthermore, defining for $\beta \in \mathbb{R}$,

$$\mathcal{M}_\beta := \left\{ b \in \overline{T(\lambda_0)} \text{ such that } \|b\|_\beta^2 := \sum_{j=1}^{\infty} \lambda_j^{-2\beta} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 < \infty \right\},$$

it is well known (see, e.g., Carrasco, Florens, and Renault, 2007) that

$$\overline{\mathcal{R}(S^*)} \equiv \mathcal{M}_0 = \left\{ b \in \overline{T(\lambda_0)} \text{ such that } \sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 < \infty \right\},$$

whereas

$$\mathcal{R}(S^*) \equiv \mathcal{M}_1 = \left\{ b \in \overline{T(\lambda_0)} \text{ such that } \sum_{j=1}^{\infty} \lambda_j^{-2} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 < \infty \right\}.$$

The following result gives sufficient, and in some cases necessary, conditions for regular and irregular identification in terms of $\|r_\phi\|_\beta$.

THEOREM 4.4. *Let Assumption 1 hold. Then, (i) $\phi(\lambda_0)$ is regularly identified iff $\|r_\phi\|_1 < \infty$, and in that case, $I_{\phi,1} = \|r_\phi\|_1^{-2}$; (ii) If $\|r_\phi\|_1 = \infty$ but $\|r_\phi\|_\beta < \infty$, for some $0 < \beta < 1$, then*

$$\inf_{b \in \mathcal{B}_\phi, ||b||_{\mathbf{H}} \leq 1} \frac{||Sb||^2}{[\dot{\phi}(b)]^{2\rho}} \geq \frac{1}{\|r_\phi\|_\beta^{2\rho}} > 0,$$

for $\rho = 1/\beta$, so $\phi(\lambda_0)$ is locally irregularly identified.

In part (i), it is possible to give another expression for the efficiency bound as a variance. The condition $\|r_\phi\|_1 < \infty$ implies there exists $i_\phi \in L_2$ such that $S^* i_\phi = r_\phi$ and $I_{\phi,1}^{-1} = \mathbb{E}[i_\phi^2(Z_i)]$. The efficiency bound is then $\mathbb{E}[i_\phi^2(Z_i)]$. It is known that in many cases, bounds on $\|r_\phi\|_\beta$ correspond to imposing smoothness conditions on r_ϕ (see Kress, 1999, Chapter 8). The smaller the β, the more irregular the identification. Hence, in these cases, one can index the level of irregularity by

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

the level of smoothness of the influence function r_ϕ. Related conditions have been extensively used in the literature of ill-posed inverse problems in statistics and econometrics as "source conditions" (see, e.g., Carrasco et al., 2007; Chen and Reiss, 2011). Our results show that the source conditions are sufficient for semiparametric identification and necessary for certain rates of convergence (cf. Theorems 4.4 and 4.3, respectively).

Example 1 (Cont.). When $\lambda_0 \notin L_2$, irregular identification holds in this example, and the linear approximation functional $\dot{\phi}(b) = \langle b, 2\lambda_0 \rangle$ for the average density becomes discontinuous in L_2. Furthermore, the score operator is not compact. Nevertheless, the proof of Theorem 4.4 suggests a general method of proof that is also valid in this setting. Consider first the average functional with infinite variance, i.e., $\phi(\lambda_0) = \mathbb{E}_{\lambda_0}[w(Z_i)]$ for a known function w satisfying $\mathbb{E}_{\lambda_0}\left[w^2(Z_i)\right] = \infty$. Assume $w \in L_q$, for some $q, 1 \leq q < 2$. Then, the Hölder inequality with $p = q/(q-1) > 2$ yields

$$[\langle b, w \rangle]^2 \leq \|b\|_p^2 \|w\|_q^2.$$

Furthermore, if

$$\|b\|_p \leq \|b\|^\beta \|b\|_r^{1-\beta}, \tag{8}$$

for some $r > p$ and $0 < \beta < 1$, then

$$\inf_{b \in \mathcal{B}_\phi, ||b||_r \leq 1} \frac{||b||^2}{|\langle b, w \rangle|^{2\rho}} \geq \frac{1}{\|w\|_q^{2\rho}} > 0, \tag{9}$$

with $\rho = 1/\beta$. Thus, a positive generalized Fisher information can be obtained for a tangent space satisfying the inequalities in (8). I note that such interpolation inequalities are available in great generality (see, e.g., Adams and Fournier, 2003, p. 27). Note the trade-off between the regularity of w and that of the tangent space for b: a lower regularity for w (a smaller q), requires a larger regularity for the tangent space (larger p, and hence larger r). A study of this trade-off and the conditions under which equality can be obtained in (9) will be investigated in future research. ▲

5. SEMIPARAMETRIC MODELS

This section studies the important class of semiparametric models, where $\mathcal{P} = \{\mathbb{P}_{\theta,\eta} : \theta \in \Theta, \eta \in H\}$. The parameter space $\Lambda = \{(\theta, \eta) : \theta \in \Theta, \eta \in H\}$ is a subset of a Hilbert space $\mathbf{H} = \mathbb{R}^p \times \mathcal{H}$. Define $\langle (\theta_1, \eta_1), (\theta_2, \eta_2) \rangle_{\mathbf{H}} := \theta_1' \theta_2 + \langle \eta_1, \eta_2 \rangle_{\mathcal{H}}$. For semiparametric models, the score operator, defined as in (2), has the representation (by the chain rule)

$$S\left(b_\theta, b_\eta\right) = \dot{l}_\theta' b_\theta + \dot{l}_\eta b_\eta, \; b = (b_\theta, b_\eta) \in T(\lambda_0) \subseteq \mathbf{H}, \tag{10}$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

where $\dot{l}_\theta \in L_2^p$ is the ordinary score function of θ and $\dot{l}_\eta$ is a continuous linear operator from $T(\eta_0) \subset \mathcal{H}$ to L_2. Let $\tilde{l}_\theta := \dot{l}_\theta - \Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}}\dot{l}_\theta$ be the so-called efficient score function for θ, where $\Pi_{\overline{V}}$ denotes the orthogonal projection operator onto $\overline{V}$. The efficient Fisher information matrix for θ is $\tilde{I}_\theta := \mathbb{E}\left[\tilde{l}_\theta \tilde{l}_\theta'\right]$. The following result provides a characterization of the main identification condition for the finite-dimensional parameter of a semiparametric model. For simplicity, I consider the case $p = 1$, and the extension to $p > 1$ follows from an application of the result to the functionals $\phi(\lambda) = \alpha'\theta$, for $\alpha \in \mathbb{R}^p$. Section 1.1 in the online Supplemental Appendix provides a parallel result for linear continuous functionals of the nuisance parameter, allowing for θ to be infinite-dimensional and possibly unidentified. The next proposition appears to be a new characterization of the main condition for local identification in semiparametric models.

PROPOSITION 5.1. *For the functional* $\phi(\lambda) = \theta \in \mathbb{R}$*:* $\mathcal{N}(I_{\lambda_0}) \subset \mathcal{N}(\dot{\phi})$ *holds iff (i)* $\dot{l}_\theta \notin \overline{\mathcal{R}(\dot{l}_\eta)}$ *(positive information* $\tilde{I}_\theta > 0$*) or (ii)* $\dot{l}_\theta \in \overline{\mathcal{R}(\dot{l}_\eta)} \setminus \mathcal{R}(\dot{l}_\eta)$ *(zero information* $\tilde{I}_\theta = 0$*).*

In the remainder of this section, I extend some of the previous results to a class of models that are nonlinear in the parameter of interest but linear in nuisance parameters.

Assumption 2. For some $\rho \geq 1$ and for all $\varepsilon > 0$, there exists $\delta > 0$ and a continuous linear operator S such that, $\left\| \left(f_\lambda - f_{\lambda_0}\right)/f_{\lambda_0} - S(\lambda - \lambda_0) \right\| < \varepsilon \left|\theta - \theta_0\right|^\rho$, for all $\lambda = (\theta, \eta) \in \mathcal{B}_\delta(\lambda_0)$.

Assumption 2 is a mean squared differentiability condition with a Lipschitz property on the derivative. It generally holds for models that are nonlinear and smooth in the parameter of interest θ, but linear in the nuisance parameters. Examples of models satisfying Assumption 2 include, among others, structural models of unemployment duration in Heckman and Singer (1984a, b); linear and nonlinear panel data models with fixed effects (see, e.g., Bonhomme, 2012); incomplete and complete games with multiple equilibria (see, e.g., Bajari et al., 2011); semiparametric measurement error models (see, e.g., Hu and Schennach, 2008); dynamic discrete choice models (see, e.g., Hu and Shum, 2012); and binary discrete choice models with single and multiple agents (see, e.g., Chamberlain, 1986, and more recently, Khan and Nekipelov 2018). Importantly, Assumption 2 allows the nonparametric parameter λ_0 to be unidentified and the parameter θ_0 to be locally irregularly identified, as it occurs in many of the aforementioned applications. The latter feature differentiates our analysis from Chen et al.'s (2014) setting. In most cases, $1 \leq \rho \leq 2$ in Assumption 2, which will limit the degree of irregularity permitted in identifying θ_0, but functionals of the nuisance parameter are allowed to have arbitrary degrees of irregularity, which can be important to accommodate many economic applications with smooth densities (e.g., Heckman and Singer, 1984a, b).

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

The generalized Fisher information for θ is

$$I_{\theta,\rho} = \inf_{b \in \mathcal{B}_\theta} \frac{||Sb||^2}{|\theta - \theta_0|^{2\rho}},$$

where $\mathcal{B}_\theta := \{b \in \overline{T(\lambda_0)} : b = (\theta - \theta_0, b_\eta), \theta \neq \theta_0, |\theta - \theta_0| \leq 1\}$. It is straightforward to show that $\tilde{I}_\theta = I_{\theta,1}$. The next theorem extends Theorem 4.2 to the nonlinear setting of Assumption 2.

THEOREM 5.1. *Let Assumption 2 hold. If $I_{\phi,\rho} > 0$, for some $\rho, 1 \leq \rho < \infty$, then θ_0 is locally identified: regularly if $\rho = 1$ and irregularly if $\rho > 1$ and $\tilde{I}_\theta = 0$.*

Theorem 5.1 extends Theorem 7 in Chen et al. (2014) to the semiparametric irregular case $\rho > 1$. If the generalized information $I_{\theta,\rho}$ is zero, Theorem 4.3 implies impossibility results on rates of convergence. Assumption 2 facilitates the verification of the conditions for Theorem 4.3 to hold. To see this, consider a path $\lambda_t \in \Lambda$ passing through λ_0 such that $\|S(\lambda_t - \lambda_0)\| = o(t^\rho)$ and $|\theta_t - \theta_0| = Ct$. For such a path, Assumption 2 yields the conditions of Theorem 4.3. Nevertheless, Assumption 2 is not necessary for Theorem 4.3 to hold.

6. EXAMPLES

6.1. Willingness to Pay

The observed data are $Z_i = (Y_i, V_i, X_i)$, where $Y_i = 1(W_i \leq V_i), W_i$ is the unobserved WTP of individual i, V_i is a continuous observed random variable with known and absolutely continuous conditional distribution $F_{V/X}$, and X_i a d-dimensional vector of covariates. The support of W is $\mathcal{S}_W := [0, w_{\max}]$ and that of V is $\mathcal{S}_V := [0, v_{\max}]$, with $0 < w_{\max}, v_{\max} \leq \infty$. It is assumed that W_i and V_i are conditionally independent given X_i. Assume $G_0(v,x) = \mathbb{P}[W_i \leq v | X_i = x]$ is differentiable at v, for each x, and define $\lambda_0(v,x) = \partial G_0(v,x)/\partial v$. The conditional density λ_0 belongs to the parameter space Λ, a subset of Lebesgue densities on $\mathcal{S}_W \times \mathcal{S}_X$. Define the measure $\mu(\{0\} \times B) = \mu(\{1\} \times B) = \mu_{VX}(B)$, where B is a Borel set of $\mathbb{R}^{d+1}$ and $\mu_{VX}(B)$ is the probability measure for (V,X). The density of Z_i wrt μ is

$$f_{\lambda_0}(y,v,x) = [G_0(v,x)]^y [1 - G_0(v,x)]^{1-y}. \tag{11}$$

Suppose one is interested in the identification of the moment functional $\phi(\lambda_0) = \mathbb{E}[r(W_i, X)]$, as in Lewbel et al. (2011). I consider first this case, and then analyze identification of the median WTP,

$$\phi(\lambda_0) = Median(W) = \inf\{w \in \mathcal{S}_W : G_0(w) \geq 1/2\},$$

where $G_0(w) = \mathbb{P}[W_i \leq w]$ is the cdf of W_i.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Following Remark 1(ii), the score operator is defined on $L_2^0(\pi)$, where π denotes the probability measure for (W, X_i) with conditional cdf $G_0(v,x)$. From (11), the score operator is given for $y \in \{0,1\}, v \in \mathcal{S}_V$, and $x \in \mathcal{S}_X$ by

$$S(b) = \frac{1}{f_{\lambda_0}(z)} 1(f_{\lambda_0} > 0)\,[2y-1] \int_0^v b(w,x)\lambda_0(w,x)\,dw. \tag{12}$$

By the Fundamental Theorem of Calculus, $\mathcal{N}(S) = \{b \in T(\lambda_0) \subset L_2^0(\pi) : b(w,x) = 0\ \pi\text{-a.s. on } 0 \le w \le v_{\max}\}$. If $v_{\max} \ge w_{\max}$, then $\mathcal{N}(S) = \{0\}$ and all linear functionals of λ_0 are identified by Proposition 4.1.

If $v_{\max} < w_{\max}$, identification is more complicated and depends on the functional. For the functional $\phi(\lambda_0) = \mathbb{E}[r(W_i, X_i)]$,

$$\begin{aligned}\dot{\phi}(b) :=& \phi(\lambda_0 + b\lambda_0) - \phi(\lambda_0)\\ =& \int_{\mathcal{S}_W \times \mathcal{S}_X} r(w,x) b(w,x) \lambda_0(w,x)\, d\mu_X dw\\ =& \langle r, b\rangle_{\mathbf{H}},\end{aligned}$$

where $\mathbf{H} = L_2(\pi)$ and μ_X is the measure of X_i. The unique Riesz representer of $\dot{\phi}$ in $\overline{T(\lambda_0)}$ is given by

$$r_\phi = \Pi_{\overline{T(\lambda_0)}}\, r.$$

A sufficient condition for identification (cf. Proposition 4.1) is

$$\int_{v_{\max}}^{w_{\max}} \int_{\mathcal{S}_X} r_\phi(w,x) b(w,x) \lambda_0(w,x)\, d\mu_X dw = 0,$$

for all $b \in \overline{T(\lambda_0)} \subset L_2^0(\pi)$. Taking $b(w,x) = r_\phi(w,x)$ leads to the following result.

PROPOSITION 6.1. *If $v_{\max} < w_{\max}$, a sufficient condition for identification of $\phi(\lambda_0) = \mathbb{E}[r(W_i, X_i)]$ with $r \in L_2(\pi)$ is*

$$r_\phi(w,x) = 0\ \pi\text{-a.s. on } v_{\max} < w \le w_{\max}.$$

Consider the general case of a nonparametric model for WTP, so $\overline{T(\lambda_0)} = L_2^0(\pi)$. In this case,

$$r_\phi = r - \phi(\lambda_0).$$

To investigate the degree of identification, whether irregular or regular, let us compute the adjoint score operator. First, note by conditional independence

$$S(b) = \frac{1}{f_{\lambda_0}(z)} 1(f_{\lambda_0} > 0)\,[2y-1]\,\mathbb{E}[\,1(W \le V) b(W, X_i) |\, V = v, X_i = x].$$

Considering separately the case $y = 1$ and $y = 0$, we see that

$$S(b) = \mathbb{E}[\,b(W, X_i) |\, Y_i = y, V = v, X_i = x]\, 1(f_{\lambda_0} > 0).$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

It is then well known that the adjoint operator of a conditional mean operator is

$$S^*g(w,x) = \mathbb{E}[\,g(Y_i,V,X_i)|\,W=w,X_i=x],$$

which after some simple algebra gives

$$S^*g(w,x) = \int_0^{v_{\max}} 1(w \le v)g(1,v,x)f_{V/X=x}(v)dv + \int 1(w>v)g(0,v,x)f_{V/X=x}(v)dv,$$

where $f_{V/X=x}(v)$ is the known conditional Lebesgue density of V_i given $X_i = x$. A direct consequence of this representation of S^* and the Fundamental Theorem of Calculus is that $S^*g(w,x)$ is absolutely continuous in w, for each x. This result can be used to characterize when moment functionals such as $\phi(\lambda_0) = \mathbb{E}[r(W_i,X_i)]$ are regularly identified. The regularity condition $r = S^*g$, for some $g \in L_2$, implies by the Fundamental Theorem of Calculus

$$\frac{\partial r(w,x)}{\partial w} = (g(0,w,x) - g(1,w,x))f_{V/X=x}(w),$$

which has a solution

$$g(y,v,x) = [1-2y]\frac{\partial r(v,x)}{\partial v}\frac{1}{f_{V/X=x}(v)}1(f_{V/X=x}(v)>0) \tag{13}$$

in L_2, provided

$$\int_0^{v_{\max}}\int\left[\frac{\partial r(v,x)}{\partial v}\right]^2\frac{1}{f_{V/X=x}(v)}d\mu_X(x)dv < \infty. \tag{14}$$

The next result summarizes the findings on moments.

PROPOSITION 6.2. *Assume* $\overline{T(\lambda_0)} = L_2^0(\pi)$ *and* $v_{\max} \ge w_{\max}$. *Let* $r \in L_2(\pi)$ *be a moment function such that* $r(w,x)$ *is differentiable at* w, *for each* x. *Then, (14) is necessary and sufficient for regular identification of* $\phi(\lambda_0) = \mathbb{E}[r(W_i,X)]$.

This result has interesting implications for known results in the literature. First, the regular estimator that results from the moment representation based on the solution (13)

$$\begin{aligned}\phi(\lambda_0) &= \mathbb{E}[g(Y_i,V_i,X_i)]\\ &= \mathbb{E}\left[[1-2Y_i]\frac{\partial r(V_i,X_i)}{\partial v}\frac{1}{f_{V/X}(V_i)}\right]\end{aligned}$$

is related to (but different from) the estimator proposed in Lewbel (1997) without covariates, which is given by

$$\mathbb{E}\left[[1(V_i \ge 0) - Y_i]\frac{\partial r(V_i)}{\partial v}\frac{1}{f_V(V_i)}\right].$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

The arguments above then show that the sufficient finite variance condition derived in Lewbel (1997) for asymptotic normality of his estimator turns out to be also necessary.

When applied to the mean of W, i.e., $r(w,x) = w$, which is a leading example in Khan and Tamer (2003), the necessary and sufficient condition for regular identification becomes simply

$$\int_0^{v_{\max}} \int \frac{1}{f_{V/X=x}(v)} d\mu_X(x)dv < \infty. \tag{15}$$

If supports of W and V_i are unbounded, i.e., $w_{\max} = v_{\max} = \infty$, so $f_{V/X=x}(u)$ vanishes in the tails, the last condition does not hold, which gives Khan and Tamer's (2010) irregularity result using a different method of proof (they compute least favorable distributions and Fisher information). Note that irregularity can also happen with bounded supports, depending on the density $f_{V/X=x}(v)$, and that the condition is necessary (these results were not considered in Khan and Tamer (2003) for the mean $r(w,x) = w$).

I now turn to the more difficult problem of semiparametric identification of the median WTP, $\phi(\lambda_0) = Median(W)$. This functional is nonlinear, and hence, the results of Theorem 4.1 are not readily applicable. For local identification, we require the following condition.

Assumption 3. (i) $v_{\max} > \phi(\lambda_0)$; and (ii) $G_0(v)$ is continuous on $\mathcal{S}_V$ and differentiable in a neighborhood of $v = \phi(\lambda_0)$ with a positive derivative $\mathbb{E}[\lambda_0(v, X_i)]$.

Assumption 3(i) is a support condition that is weaker than that for identification of moments of WTP. Assumption 3(ii) is typical in quantile identification. Under Assumption 3, the median WTP has an influence function

$$r_\phi(w,x) = \frac{-\{1(w < \phi(\lambda_0)) - 0.5\}}{\mathbb{E}[\lambda_0(\phi(\lambda_0), X_i)]}.$$

As the result above suggests, the discontinuity of the influence function implies irregular identification. The next result formalizes this finding.

PROPOSITION 6.3. *Let Assumption 3 hold, and assume* $\overline{T(\lambda_0)} = L_2^0(\pi)$. *Then,* $\phi(\lambda_0) = Median(W)$ *is irregularly identified.*

6.2. Asset Pricing Euler Equation with Measurement Error

The goal of this example is to provide primitive conditions based on the results of this paper for identification of the discount factor θ_0 and measures of risk aversion. These are important parameters in these models. For example, the discount factor is a key determinant of an individual's intertemporal decisions such as asset accumulation (Venti and Wise, 1998; Samwick, 2006), labor supply decisions (MaCurdy, 1981) and job search (MaCurdy, 1981).

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

6.2.1. *Identification of the Discount Factor.* The Euler equation with measurement error is a nonlinear conditional moment restriction model. The first step in our analysis is to parametrize the model in a way that makes it amenable to the results of this paper. To that end, I consider the following assumption. Recall $C_t = m(C_t^*, \varepsilon_t)$.

Assumption 4. (i) C_{t+1}^* is independent of C_t, conditional on C_{t+1}, and measurable wrt $\sigma(C_{t+1}, \varepsilon_{t+1})$; (ii) $(\varepsilon_{t+1}, \varepsilon_t)$ is independent of R_{t+1}, given (C_{t+1}, C_t); and (iii) the distribution of C_t^* conditional on C_t does not depend on t.

Assumption 4(i) can be relaxed to: C_{t+1}^* is independent of C_{t-1}, conditional on (C_{t+1}, C_t), at the cost of increasing the dimension of the arguments in the nonparametric component given below. If m is monotone in C_t^*, then 4(i) can be written in terms of ε_{t+1}. Assumption 4(i–iii) is less restrictive than typical assumptions considered in the literature, which assume, in addition to functional form assumptions on m, that ε_t is independent of "everything" (see, for example, Altonji and Siow, 1987; Runkle, 1991; Dynam, 2000; Alan et al., 2009).

For the sake of exposition, I consider the case without household characteristics $X_{t,i}$. The presence of $X_{t,i}$ in $\mathcal{F}_{t,i}$ adds additional moment restrictions, so it is simpler for identification. All the arguments below can be easily adapted to the presence of $X_{t,i}$. It is also straightforward to extend the identification results to models with more than one asset, habit formation or other observable variables in the marginal utility.

Assumption 4 ensures the following parametrization in terms of observables (C_{t+1}, R_{t+1}, C_t) :

$$\begin{aligned}
&\mathbb{E}\left[\theta_0 \dot{u}(C_{t+1}^*) R_{t+1} - \dot{u}(C_t^*) \middle| C_t\right] \\
&\quad = \mathbb{E}\left[\theta_0 \dot{u}(C_{t+1}^*) \mathbb{E}[R_{t+1} | C_{t+1}, C_t, \varepsilon_{t+1}, \varepsilon_t] - \dot{u}(C_t^*) \middle| C_t\right] \\
&\quad = \mathbb{E}\left[\theta_0 \dot{u}(C_{t+1}^*) \mathbb{E}[R_{t+1} | C_{t+1}, C_t] - \dot{u}(C_t^*) \middle| C_t\right], \\
&\quad = \mathbb{E}\left[\theta_0 \mathbb{E}\left[\dot{u}(C_{t+1}^*) \middle| C_{t+1}\right] \mathbb{E}[R_{t+1} | C_{t+1}, C_t] - \dot{u}(C_t^*) \middle| C_t\right] \\
&\quad = \mathbb{E}[\theta_0 \eta_0(C_{t+1}) R_{t+1} - \eta_0(C_t) | C_t],
\end{aligned}$$

where $\eta_0(C_{t+1}) = \mathbb{E}\left[\dot{u}(C_{t+1}^*) \middle| C_{t+1}\right]$ and the first equality uses Assumption 4(i), the second Assumption 4(ii), the third Assumption 4(i), and the fourth Assumption 4(iii). This new parametrization is a nonlinear conditional moment indexed by the discount factor θ_0 and the projected marginal utility η_0. Let μ be the probability measure of C_t. Then, I denote $\lambda_0 = (\theta_0, \eta_0) \in \Lambda = \Theta \times H, \Theta \subset (0, 1)$ and $H \subset L_2(\mu)$.

The following condition guarantees that the conditional mean operator

$$A\eta(c) = \mathbb{E}[\eta(C_{t+1}) R_{t+1} | C_t = c] \tag{16}$$

is well defined and compact when viewed as $A : L_2(\mu) \to L_2(\mu)$. This is a standard assumption in the literature (see, e.g., Carrasco et al., 2007). Let $g(C_{t+1}, C_t)$ be the joint Lebesgue density of (C_{t+1}, C_t), and let $f_{t+1}(C_{t+1})$ and $f_t(C_t)$ denote its marginals, respectively.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Assumption 4. (iv) $0 < \mathbb{E}\left[R_{t+1}^2 \{g(C_{t+1}, C_t)/f_t(C_{t+1})f_t(C_t)\}\right] < \infty$.
Define, for $\lambda = (\theta, \eta) \in \Lambda = \Theta \times H$,

$$M_\lambda = \theta A\eta(c) - \eta(c).$$

The identification results of Section 5 can be adapted to this moment model replacing $\left(f_\lambda - f_{\lambda_0}\right)/f_{\lambda_0}$ by $M_{\lambda_0+b} - M_{\lambda_0}$ and the norm $\|\cdot\|$ by $\|\cdot\|_{2,\mu}$. I then proceed to verify Assumption 2 in this example. It is straightforward to show that for all $b = (b_\theta, b_\eta) \in \mathbb{R} \times L_2(\mu)$, with $b_\theta = \theta - \theta_0$,

$$\left\|M_{\lambda_0+b} - M_{\lambda_0} - S(b)\right\|_{2,\mu} \leq \left\|\theta_0 A b_\eta(c) - b_\eta(c)\right\|_{2,\mu} |b_\theta|, \tag{17}$$

where

$$S\left(b_\theta, b_\eta\right) = \dot{l}_\theta b_\theta + \dot{l}_\eta b_\eta,$$

$$\dot{l}_\theta b_\theta = b_\theta A\eta_0 \text{ and } \dot{l}_\eta b_\eta = \theta_0 A b_\eta(c) - b_\eta(c).$$

Since A is bounded, for each $\varepsilon > 0$, one can make $\left\|\theta_0 A b_\eta(c) - b_\eta(c)\right\|_{2,\mu} < \varepsilon$ by choosing δ small enough and $\left\|b_\eta\right\|_{2,\mu} < \delta$. Thus, Assumption 2 holds with $\rho = 1$.

It follows from the previous parametrization and Assumption 4 that local identification of the discount factor is regular.[4] Formally, Theorem 3.2 in Kress (1999, p. 29) implies that the range of $\dot{l}_\eta, \mathcal{R}(\dot{l}_\eta)$, is closed. It follows from Rudin (1973, Theorem 4.14) that $\mathcal{R}(\dot{l}_\eta^*)$ is also closed, and by the expression above, $\mathcal{R}(S^*)$ is also closed. Then, by the results of this paper, all locally identified linear continuous functionals of $\lambda_0 = (\theta_0, \eta_0)$ are regularly locally identified. Note that this does not apply to functionals of the marginal utility $\dot{u}_0$.

Positive information for the discount factor θ_0 holds iff $(1, 0) \in \mathcal{R}(S^*)$, which means that there exists $g \in L_2(\mu)$ such that

$$\langle A\eta_0, g\rangle = 1, \dot{l}_\eta^* g = 0. \tag{18}$$

Since $\dot{l}_\eta^* g = \theta_0 A^* g(c) - g(c)$, the equation $\dot{l}_\eta^* g = 0$ means that g is an eigenfunction of A^* with eigenvalue θ_0^{-1}. Such an eigenfunction exists, because eigenvalues of A^* are complex conjugates of those of A and θ_0 is real-valued. Then, a sufficient condition for local identification of the discount factor is that for such eigenfunction, say $g_0 : \langle \eta_0, g_0\rangle \neq 0$. Note that $\langle \eta_0, g_0\rangle \neq 0$ and $\theta_0 > 0$ guarantee (18) by choosing $g = cg_0$ with $c = 1/\langle A\eta_0, g_0\rangle$ (since multiples of eigenfunctions are eigenfunctions). The discussion is summarized in the following result.

PROPOSITION 6.4. *Let Assumption 4 hold and assume* $\langle \eta_0, g_0\rangle \neq 0$ *for* g_0 *one of the eigenfunctions of* A^* *corresponding to the eigenvalue* θ_0^{-1}. *Then,* θ_0 *is locally regularly identified.*

[4]The actual effective score operator of the model is proportional to S, with a proportionality "constant" given by the inverse of the conditional variance $\mathbb{E}\left[V_{t+1}^2 \middle| C_t = c\right]$, where $V_{t+1} = \theta_0 \eta_0(C_{t+1}) R_{t+1} - \eta_0(C_t)$. This conditional variance is assumed to be bounded and bounded away from zero. Thus, statements related to regularity can be given in terms of S.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

The condition $\langle \eta_0, g_0 \rangle \neq 0$ is mild, and holds, for example, when η_0 and g_0 are positive. Escanciano and Hoderlein (2010) present primitive conditions for nonparametric identification of positive η_0 and g_0 based on Perron–Frobenius theory. Chen et al. (2014) and Escanciano et al. (2015) also use Perron–Frobenius to obtain identification of related but different Euler equation models. See also Hansen and Scheinkman (2009) and Christensen (2017) for other applications of Perron–Frobenius theory. These identification results are nonparametric and for models without measurement error. In contrast, a simple *semiparametric* identification condition for the discount factor is presented here, in a model with measurement error, namely $\langle \eta_0, g_0 \rangle \neq 0$. Proposition 6.4 thus shows that regular local identification of the discount factor holds under rather general conditions on the measurement error mechanism (m and the distribution of ε_t are nonparametric and unidentified under our conditions).

An important empirical literature has provided estimation and inference results on Euler equations accounting for measurement error. Papers within this literature use functional form assumptions for utilities and for the measurement error mechanism. The identification result of Proposition 6.4 opens the door for more robust empirical strategies for inference on the discount factor in microeconomic applications based on the Euler equation. For example, Altonji and Siow (1987), Runkle (1991), Dynam (2000), and Alan et al. (2009) assume parametric marginal utilities and $m(C_t^*, \varepsilon_t) = C_t^* \varepsilon_t$, i.e.,

$$C_t = C_t^* \varepsilon_t, \tag{19}$$

with ε_t independent of everything and, in some cases, assumed to be lognormally distributed. The identification result above shows that regular local identification of the discount factor follows under more general assumptions than previously recognized, including situations where the marginal utility, the measurement equation, and other nonparametric parameters are not identified. This point illustrates the concept of semiparametric identification emphasized in this paper.

6.2.2. *Identification of Average Risk Aversion.* The Average Arrow–Pratt coefficient of AARA parameter is given by

$$\chi(\dot{u}_0) = \mathbb{E}\left[\frac{-\partial \dot{u}_0(C_t^*)/\partial C_t^*}{\dot{u}_0(C_t^*)}\right].$$

The following conditions guarantee that this parameter is well defined, and satisfies some properties given below. Let μ^* denote the probability measure of C_t^*, with density $f_t^*(\cdot)$, and assume the parameter space for marginal utilities $\dot{U}$ satisfies, for a small positive number ϵ,

$$\dot{U} \subset \left\{\dot{u} \in L_2(\mu^*) : \dot{u}(c^*) \geq \epsilon > 0 \text{ and } \int \dot{u}^2 d\mu^* = 1\right\}.$$

The scale normalization in $\dot{U}$ is necessary to achieve identification in nonparametric Euler equations.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Assumption 4. (v) The functions $\log(\dot{u}_0(\cdot))$ and $\log(f_t^*(\cdot))$ are continuously differentiable on the convex support of C_t^* (possibly unbounded) and these functions and their derivatives are in $L_2(\mu^*)$. The true marginal utility satisfies $\dot{u}_0 \in \dot{U}$.

By Assumption 4(v) and integration by parts,

$$\chi(\dot{u}_0) = \mathbb{E}\left[\log\left(\dot{u}_0(C_t^*)\right) d(C_t^*)\right],$$

where

$$d(c^*) \equiv \frac{\partial f_t^*(c^*)}{\partial c^*} \frac{1}{f_t^*(c^*)}. \tag{20}$$

The functional $\chi(\dot{u}_0)$, although nonlinear, is concave and differentiable, with pathwise derivative

$$\dot{\chi}(b) = \mathbb{E}\left[b(C_t^*) \frac{d(C_t^*)}{\dot{u}_0(C_t^*)}\right].$$

Thus, the AARA parameter as a functional on $L_2(\mu^*)$ has the Riesz representer

$$r_\chi(C_t^*) = \frac{d(C_t^*)}{\dot{u}_0(C_t^*)}. \tag{21}$$

To link the marginal utility with the projected marginal utility, we need to define the conditional mean operator $L : L_2(\mu^*) \to L_2(\mu)$,

$$L\dot{u}(C_t) = \mathbb{E}\left[\dot{u}(C_t^*) \middle| C_t\right],$$

which has an adjoint operator $L^* : L_2(\mu) \to L_2(\mu^*)$ given by

$$L^* w(C_t) = \mathbb{E}\left[w(C_t) | C_t^*\right].$$

Applying the results of this paper, one obtains the following identification result for the AARA.

PROPOSITION 6.5. *Let Assumption 4 hold and assume θ_0 is locally identified. Then, the following condition is sufficient for regular local identification of the AARA: there exists $g \in L_2(\mu)$ such that, with r_χ given in (21) and $\dot{l}_\eta^* g = \theta_0 A^* g(c) - g(c)$,*

$$r_\chi(C_t^*) = \mathbb{E}\left[\dot{l}_\eta^* g(C_t) \middle| C_t^*\right]. \tag{22}$$

It is convenient to decompose (22) into two parts: (i) existence of $w \in L_2(\mu)$ such that

$$r_\chi(C_t^*) = \mathbb{E}\left[w(C_t) | C_t^*\right]; \tag{23}$$

and (ii) conditions that guarantee that such w belongs to $\mathcal{R}(\dot{l}_\eta^*)$. I provide simple primitive conditions for (23) to hold in the multiplicative measurement error model

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

(19), when the error density is known and given by f_ε, e.g., lognormal as in, e.g., Alan et al. (2009). In this model, the regularity condition (23) is,

$$r_\chi(c^*) = \int f_\varepsilon(c/c^*)w(c)dc.$$

The following lemma provides a sufficient condition for existence of $w \in L_2(\mu)$ satisfying this equation. Let $L_1(\mathbb{R})$ and $L_2(\mathbb{R})$ denote the set of integrable and squared integrable functions, respectively. For $f \in L_1(\mathbb{R})$, define the Fourier transform $\hat{f} = (2\pi)^{-1/2}\int e^{-itx}f(t)dt$, where $i = \sqrt{-1}$. Define $K(u) = \exp(u)f_\varepsilon(\exp(u))$ and $x(\tau) = \exp(-\tau)r_\chi(\exp(\tau))$.

LEMMA 6.1. *If $K(u)$ is symmetric in $u, x \in L_2(\mathbb{R})$ and $\hat{x}/\hat{K} \in L_2(\mathbb{R})$, then there exists a solution $w \in L_2(\mu)$ of (23). Moreover, a solution is given by*

$$w(c) = \frac{1}{2\pi}\,\mathrm{Re}\int e^{it\log(c)}\frac{\hat{x}(t)}{\hat{K}(t)}dt,$$

where Re *denotes the real part.*

The symmetry condition on K is satisfied by the lognormal distribution used in the empirical literature. I now provide primitive conditions for $w \in \mathcal{R}(\dot{l}^*_\eta)$. Note that if $w \notin \mathcal{R}(\dot{l}^*_\eta)$, the AARA is not identified. By duality, $w \in \mathcal{R}(\dot{l}^*_\eta) = \mathcal{N}(\dot{l}_\eta)^\perp$ has a simple interpretation: w is orthogonal to all projected marginal utilities solving the Euler equation, i.e.,

$$\mathbb{E}[w(C_t)\eta(C_t)] = 0 \text{ for all } \eta \text{ such that } \theta_0 A\eta = \eta. \tag{24}$$

By compactness of A, the space of such $\eta's$ is finite-dimensional (see Kress, 1999), which means that (24) can be tested. Importantly, (24) holds for η_0 under (23), since by iterated expectations,

$$\begin{aligned}\mathbb{E}[w(C_t)\eta_0(C_t)] &= \mathbb{E}\left[r_\chi(C_t^*)\eta_0(C_t)\right]\\ &= \mathbb{E}\left[r_\chi(C_t^*)\dot{u}_0(C_t^*)\right]\\ &= \mathbb{E}\left[d(C_t^*)\right]\\ &= 0.\end{aligned}$$

A primitive condition for identification of η_0 is $r(C_{t+1}, C_t) > 0$ and $g(C_{t+1}, C_t) > 0$, where $r(C_{t+1}, C_t) = \mathbb{E}[R_{t+1}|\, C_{t+1}, C_t]$. Thus, these primitive conditions and the mild integrability conditions of Lemma 6.1 imply regular identification of the AARA by virtue of Proposition 6.5.

7. CONCLUSIONS

This paper provides tools for investigating semiparametric identification, with a particular emphasis on irregular identification. First, it considers semiparametric identification for linear models and obtains necessary and sufficient conditions

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

for regular and irregular identification. I apply these results to show that the median WTP is irregularly identified in contingent valuation studies. The Euler equation application illustrates the usefulness of the characterization of regular identification and its applicability to the important class of conditional moment models. Regular identification of the discount factor and measures of risk aversion are obtained under relatively simple conditions, despite the nonlinearity of both the model and the AARA functional and the nonparametric nature of the measurement error.

The question of whether zero semiparametric information corresponds to a lack of identification is a rather delicate question, as was first pointed out in Chamberlain (1986). Indeed, I show here that irregularity corresponds mathematically to a boundary case in an infinite-dimensional space. Regular identification is, however, easier to characterize, and, under mild smoothness conditions, a positive semiparametric Fisher information for the parameter implies its local identification. When Fisher information is zero, positivity of a new generalized Fisher information introduced in this paper implies irregular identification. When the generalized Fisher information is zero, I obtain impossibility results on rates of convergence. The impossibility results on regular identification and rates apply to both linear and nonlinear models and parameters.

A number of issues deserve further study. The delicate role of nonlinearity in identification, first pointed out in Sargan (1958, 1983) and recently studied in the nonparametric identification setting by Chen et al. (2014), deserves further investigation. More generally, it will be useful to investigate primitive conditions for positive or zero generalized Fisher information in specific economic applications and, using the tools provided here, to see how these conditions translate into specific rates of convergence for estimators. Likewise, Section 1.3 in the online Supplemental Appendix provides sufficient conditions for regular and irregular semiparametric identification in nonlinear models and for nonlinear functionals. Applying these results to specific examples, and establishing connections with attainability of rates of convergence for semiparametric estimators, remain topics for future research.

APPENDIX: Proofs of Main Results

Proof of Proposition 4.1. I first show that $\mathcal{N}(I_{\lambda_0}) = \mathcal{N}(S)$. From the definition of I_{λ_0}, the implication $\mathcal{N}(S) \subset \mathcal{N}(I_{\lambda_0})$ trivially holds. The other implication follows from $\langle I_{\lambda_0} b, b\rangle_{\mathbf{H}} = \|Sb\|^2$. Having shown $\mathcal{N}(I_{\lambda_0}) = \mathcal{N}(S)$, that (4) is a sufficient condition for identification of $\phi(\lambda_0)$ follows from the definition of identification. If $f_\lambda = f_{\lambda_0}$, then $b = \lambda - \lambda_0 \in \mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$, and hence, $\phi(\lambda_0 + b) = \phi(\lambda_0)$. To prove the necessity, suppose that (4) does not hold, i.e.,

$$\mathcal{N}(S) \subsetneq \mathcal{N}(\dot{\phi}),$$

then there exists $b \in T(\lambda_0)$ such that $b \in \mathcal{N}(S)$ but $b \notin \mathcal{N}(\dot{\phi})$. This means by linearity that, for all $c \in \mathbb{R}, c \neq 0, cb \in \mathcal{N}(S)$ but $cb \notin \mathcal{N}(\dot{\phi})$. By Assumption 1(iii), there exists c

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

such that $\lambda_0 + cb \in \Lambda$, $cb \in \mathcal{N}(S)$, and $cb \notin \mathcal{N}(\dot{\phi})$. This implies that $f_{\lambda_0+cb} = f_{\lambda_0}$ but $\phi(\lambda_0 + cb) \neq \phi(\lambda_0)$. That is, $\phi(\lambda_0)$ is not identified. ■

Proof of Theorem 4.1. By Proposition 1 in Luenberger (1997, p. 52),

$$\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$$

is equivalent to

$$\mathcal{N}(\dot{\phi})^{\perp} \subset \mathcal{N}(S)^{\perp},$$

since both $\mathcal{N}(S)$ and $\mathcal{N}(\dot{\phi})$ are closed linear subspaces, and where henceforth $V^{\perp}$ denotes the orthocomplement of the subspace V. However, since

$$\dot{\phi}(b) = \langle b, r_{\phi} \rangle_{\mathbf{H}},$$

for all $b \in T(\lambda_0)$, it follows that $\mathcal{N}(\dot{\phi})^{\perp} = \left(\{r_{\phi}\}^{\perp}\right)^{\perp} = span\{r_{\phi}\}$ (by Proposition 1.5 in Luenberger (1997, p. 52)). On the other hand, by Theorem 3 in Luenberger (1997, p. 157),

$$\mathcal{N}(S)^{\perp} = \overline{\mathcal{R}(S^*)}.$$

The identification part follows from Proposition **4.1**. The qualification of regular or irregular follows from Theorem 4.1 in van der Vaart (1991), which shows that $r_{\phi} \in \mathcal{R}(S^*) \Longleftrightarrow I_{\phi} > 0$. ■

Proof of Theorem 4.2. Suppose $\phi(\lambda_0)$ is not identified. That means we can find $\lambda \in \Lambda$ such that $\phi(\lambda) \neq \phi(\lambda_0)$ and $f_{\lambda} = f_{\lambda_0}$. Then, $b = \lambda - \lambda_0 \in \mathcal{N}(S)$ but $\dot{\phi}(b) \neq 0$. Choose c sufficiently small, so that $\left|\dot{\phi}(cb)\right| \leq 1$, and hence, $cb \in \mathcal{B}_{\phi}$ and

$$I_{\phi,\rho} = \inf_{b \in \mathcal{B}_{\phi}} \frac{||Sb||^2}{[\dot{\phi}(b)]^{2\rho}} = 0,$$

contradicting the positivity of $I_{\phi,\rho}$. ■

Proof of Theorem 4.3. Fix $K > 0$ and take $t \equiv t_n = (2K/C)n^{-1/(2\rho)}$ and $\lambda_n = \lambda_{t_n}$, so that $|\phi(\lambda_n) - \phi(\lambda_0)| \geq 2Kn^{-1/(2\rho)}$ and $\left\| \left(f_{\lambda_n} - f_{\lambda_0}\right)/f_{\lambda_0} \right\|^2 < \varepsilon(2K/C)^{2\rho} n^{-1}$. Let $\mathbb{P}_n$ denote the probability measure of f_{λ_n}, and $\mathbb{P}_0$ that of f_{λ_0}, and $\mathbb{P}_n^n$ and $\mathbb{P}_0^n$ their corresponding n-fold product. Lemma 1 in LeCam (1973) and the basic inequality $\left(\sqrt{a} - \sqrt{b}\right)^2 \leq (a - b)^2/b$, for $a, b > 0$, imply that, for each $\delta > 0$, we can find $\gamma > 0$ sufficiently small such that

$$v(\mathbb{P}_n^n, \mathbb{P}_0^n) < \delta \text{ if } \left\| \left(f_{\lambda_n} - f_{\lambda_0}\right)/f_{\lambda_0} \right\|^2 < \gamma/n,$$

where $v(\mathbb{P}, \mathbb{Q})$ denotes the total variation distance between $\mathbb{P}$ and $\mathbb{Q}$. The details are given as follows. Lemma 1 in LeCam (1973) gives the inequality

$$v(\mathbb{P}_n^n, \mathbb{P}_0^n) \leq y(2 - y^2)^{1/2},$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

if $H(\mathbb{P}_n, \mathbb{P}_0) \leq y/\sqrt{n}$ and $y \leq 1$. From the basic inequality $\left(\sqrt{a} - \sqrt{b}\right)^2 \leq (a-b)^2/b$, for $a, b > 0$, we have

$$\begin{aligned} H^2(\mathbb{P}_n, \mathbb{P}_0) &= \frac{1}{2}\int \left(\sqrt{f_{\lambda_n}} - \sqrt{f_{\lambda_0}}\right)^2 d\mu \\ &\leq \frac{1}{2}\int \left[\left(f_{\lambda_n} - f_{\lambda_0}\right)^2 / f_{\lambda_0}\right] d\mu \\ &= \frac{1}{2}\left\| \left(f_{\lambda_n} - f_{\lambda_0}\right)/f_{\lambda_0} \right\|^2 \\ &< y^2/n, \end{aligned}$$

with $y^2 = \varepsilon(2K/C)^{2\rho}/2$ (where the last inequality follows from the assumptions of the theorem). Since we can always choose $\varepsilon > 0$ small enough so that $y \leq 1$, Lemma 1 in LeCam (1973) can be applied to achieve $v(\mathbb{P}_n^n, \mathbb{P}_0^n) < \delta$ by making y small enough. Next, Lemma 8 in Ishwaran (1996) implies that the rate of any estimator cannot be better than $O_P(2Kn^{-1/(2\rho)})$. Since this holds for each fixed $K > 0$, it follows that the rate of any estimator must be slower than $n^{-1/(2\rho)}$. ■

Proof of Theorem 4.4. The first part follows from Theorem 4.1(i) and the characterization of $\mathcal{R}(S^*)$ as those elements $b \in \overline{T(\lambda_0)}$ with $\|b\|_1 < \infty$. Any element $b \in \overline{T(\lambda_0)}$ has the singular value expansion (cf. Kress, 1999, Theorem 15.16)

$$b = \sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}} \varphi_j + \Pi_{\mathcal{N}(S)} b,$$

which implies under semiparametric identification of the functional

$$\dot{\phi}(b) = \sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}} \langle r_\phi, \varphi_j \rangle_{\mathbf{H}}$$

and

$$Sb = \sum_{j=1}^{\infty} \lambda_j \langle b, \varphi_j \rangle_{\mathbf{H}} \psi_j.$$

By Cauchy–Schwarz, for $b \in T(\lambda_0)$,

$$\begin{aligned} \left|\dot{\phi}(b)\right| &\leq \left(\sum_{j=1}^{\infty} \lambda_j^{-2} \langle r_\phi, \varphi_j \rangle_{\mathbf{H}}^2\right)^{1/2} \left(\sum_{j=1}^{\infty} \lambda_j^2 \langle b, \varphi_j \rangle_{\mathbf{H}}^2\right)^{1/2} \\ &= \|r_\phi\|_1 \, ||Sb||. \end{aligned}$$

This yields

$$I_{\phi,1} \geq \|r_\phi\|_1^{-2}.$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

By regular identification, there exists i_ϕ with finite variance such that $S^* i_\phi = r_\phi$, and hence, another version of Cauchy–Schwarz gives

$$\begin{aligned}\left|\dot{\phi}(b)\right| &= \left\langle Sb, i_\phi \right\rangle_{\mathbf{H}} \\ &\leq \left\| i_\phi \right\| ||Sb||.\end{aligned}$$

Indeed, it can be shown that $\left\| i_\phi \right\| = \left\| r_\phi \right\|_1$ and equality in Cauchy–Schwarz holds when Sb and i_ϕ are proportional. The formal argument requires constructing

$$b_k = \frac{\sum_{j=1}^k \lambda_j^{-1} \left\langle i_\phi, \psi_j \right\rangle \varphi_j}{\sum_{j=1}^k \left\langle i_\phi, \psi_j \right\rangle^2}.$$

Note $\dot{\phi}(\varphi_j) = \left\langle r_\phi, \varphi_j \right\rangle_{\mathbf{H}} = \lambda_j \left\langle i_\phi, \psi_j \right\rangle$ (since $S\varphi_j = \lambda_j \psi_j$) and hence

$$\left|\dot{\phi}(b_k)\right| = 1,$$

and b_k is in the set $\mathcal{B}_\phi = \left\{ b \in T(\lambda_0) : \dot{\phi}(b) \neq 0, \left|\dot{\phi}(b)\right| \leq 1 \right\}$. Moreover,

$$||sb_k||^2 = \frac{1}{\sum_{j=1}^k \left\langle i_\phi, \psi_j \right\rangle^2},$$

which can get arbitrarily close to the bound $\left\| i_\phi \right\|^{-2} = \left\| r_\phi \right\|_1^{-2}$ by choosing a large k. Thus, we conclude $I_{\phi,1} = \left\| r_\phi \right\|_1^{-2} = \left\| i_\phi \right\|^{-2}$.

For the second part, note that $\left\| r_\phi \right\|_\beta < \infty$ implies $r_\phi \in \overline{\mathcal{R}(S^*)}$, and hence identification. The condition $\left\| r_\phi \right\|_1 = \infty$ then implies zero Fisher information by the first part of the theorem. On the other hand, by the Holder inequality with $p = 1/\beta$ and $q = 1/(1-\beta)$, for any $0 < \beta < 1$, and for all b with $||b||_{\mathbf{H}} \leq 1$,

$$\begin{aligned}\left|\dot{\phi}(b)\right| &\leq \left(\sum_{j=1}^\infty \lambda_j^{-2\beta} \left\langle r_\phi, \varphi_j \right\rangle_{\mathbf{H}}^2 \right)^{1/2} \left(\sum_{j=1}^\infty \lambda_j^{2\beta} \left\langle b, \varphi_j \right\rangle_{\mathbf{H}}^2 \right)^{1/2} \\ &\leq \left\| r_\phi \right\|_\beta \left(\sum_{j=1}^\infty \lambda_j^{2\beta} \left\langle b, \varphi_j \right\rangle_{\mathbf{H}}^{2\beta} \left\langle b, \varphi_j \right\rangle_{\mathbf{H}}^{2-2\beta} \right)^{1/2} \\ &\leq \left\| r_\phi \right\|_\beta \left(\sum_{j=1}^\infty \lambda_j^2 \left\langle b, \varphi_j \right\rangle_{\mathbf{H}}^2 \right)^{\beta/2} \left(\sum_{j=1}^\infty \left\langle b, \varphi_j \right\rangle_{\mathbf{H}}^2 \right)^{(1-\beta)/2} \\ &\leq \left\| r_\phi \right\|_\beta ||Sb||^\beta.\end{aligned}$$

Thus, for $\rho = 1/\beta$,

$$\inf_{b \in \mathcal{B}_\phi, ||b||_{\mathbf{H}} \leq 1} \frac{||Sb||^2}{[\dot{\phi}(b)]^{2\rho}} \geq \frac{1}{\left\| r_\phi \right\|_\beta^{2\rho}} > 0,$$

and local identification follows. ■

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Proof of Proposition 5.1. For the functional $\phi(\lambda)=\theta$, it holds $\mathcal{N}(\dot{\phi})=\{(b_\theta,b_\eta):b_\theta=0\}$. Then, by orthogonality,

$$\begin{aligned}\mathcal{N}(S)&=\left\{(b_\theta,b_\eta):\int(\dot{l}'_\theta b_\theta+\dot{l}_\eta b_\eta)^2 d\mathbb{P}_{\theta_0,\eta_0}=0\right\}\\&=\left\{(b_\theta,b_\eta):\int(\tilde{l}'_\theta b_\theta+\Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}}\dot{l}'_\theta b_\theta+\dot{l}_\eta b_\eta)^2 d\mathbb{P}_{\theta_0,\eta_0}=0\right\}\\&=\left\{(b_\theta,b_\eta):\int\left(\tilde{l}'_\theta b_\theta\right)^2 d\mathbb{P}_{\theta_0,\eta_0}=0,\int(\Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}}\dot{l}'_\theta b_\theta+\dot{l}_\eta b_\eta)^2 d\mathbb{P}_{\theta_0,\eta_0}=0\right\}\\&=\left\{(b_\theta,b_\eta):b'_\theta\tilde{I}_\theta b_\theta=0,\Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}}\dot{l}'_\theta b_\theta=-\dot{l}_\eta b_\eta\right\}.\end{aligned}$$

Then, if $\tilde{I}_\theta>0$, then we have $\mathcal{N}(S)\subset\mathcal{N}(\dot{\phi})$. If $\tilde{I}_\theta=0$, then there are two cases: (1) $\dot{l}_\theta\in\mathcal{R}(\dot{l}_\eta)$ and (2) $\dot{l}_\theta\in\overline{\mathcal{R}(\dot{l}_\eta)}\setminus\mathcal{R}(\dot{l}_\eta)$. In case (1), the identification condition $\mathcal{N}(S)\subset\mathcal{N}(\dot{\phi})$ does not hold, as we can find $b_\theta\neq0$ such that $\dot{l}_\theta b_\theta=-\dot{l}_\eta b_\eta$ (so $(b_\theta,b_\eta)\in\mathcal{N}(S)$ but $(b_\theta,b_\eta)\notin\mathcal{N}(\dot{\phi})$). In case (2), $\mathcal{N}(S)\subset\mathcal{N}(\dot{\phi})$ holds even though there is zero information for the parameter. Thus, if $\dot{l}_\theta\in\overline{\mathcal{R}(\dot{l}_\eta)}\setminus\mathcal{R}(\dot{l}_\eta)$, we have "irregular identification." ■

Proof of Theorem 5.1. Choose $0<\varepsilon<I_{\theta,\rho}^{1/2}$, and $\delta>0$, such that, for all $\lambda\in\mathcal{B}_\delta(\lambda_0)$ with $\theta\neq\theta_0$,

$$\begin{aligned}\frac{\left\|\left(f_\lambda-f_{\lambda_0}\right)/f_{\lambda_0}-S(\lambda-\lambda_0)\right\|}{\|S(\lambda-\lambda_0)\|}&=\frac{\left\|\left(f_\lambda-f_{\lambda_0}\right)/f_{\lambda_0}-S(\lambda-\lambda_0)\right\|}{|\theta-\theta_0|^\rho}\frac{|\theta-\theta_0|^\rho}{\|S(\lambda-\lambda_0)\|}\\&\leq\varepsilon\times I_{\theta,\rho}^{-1/2}\\&<1,\end{aligned}\tag{25}$$

where we have used Assumption 2 and the definition of the generalized Fisher information. The inequality (25) implies that $\left\|\left(f_\lambda-f_{\lambda_0}\right)/f_{\lambda_0}\right\|\neq0$, or equivalently, $f_\lambda\neq f_{\lambda_0}$. That is, local identification holds. ■

Proof of Proposition 6.1. The score operator and the functional satisfy Assumption 1(i,ii,iv), since $r\in L_2(\pi)$. The condition

$$r_\phi(w,x)=0\ \pi-\text{a.s. on } v_{\max}<w\leq w_{\max}$$

then implies (4). By Proposition 4.1 and the discussion of Assumption 1 (see also Remark 2), identification follows. ■

Proof of Proposition 6.2. To show that (14) is sufficient for regular identification, note that (13) is a solution to $r=S^*g$ with $g\in L_2$. To see that $g\in L_2$, note

$$\begin{aligned}\mathbb{E}\left[g^2(Z_i)\right]&=\mathbb{E}\left[g^2(1,V_i,X_i)G_0(V_i,X_i)+g^2(0,V_i,X_i)(1-G_0(V_i,X_i))\right]\\&\leq\mathbb{E}\left[g^2(1,V_i,X_i)+g^2(0,V_i,X_i)\right]\\&\leq2\int_0^{v_{\max}}\int\left[\frac{\partial r(v,x)}{\partial v}\right]^2\frac{1}{f_{V/X=x}(v)}d\mu_X(x)dv<\infty.\end{aligned}$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

The necessity follows from the arguments prior to the statement of the Proposition, i.e., from

$$\frac{\partial r(w,x)}{\partial w} = (g(0,w,x) - g(1,w,x)) f_{V/X=x}(w),$$

and the fact that $g \in L_2$. Conclude by an application of Theorem 4.1(i). ■

Proof of Proposition 6.3. I first verify that median WTP is identified. This is implied by continuity $G_0(v)$ and the support condition $v_{\max} \geq \phi(\lambda_0)$, since by the conditional independence,

$$G_0(w) = \mathbb{E}[\mathbb{E}[Y_i | X_i, V_i = w]]$$

is identified on $\mathcal{S}_V$ and $G_0(v_{\max}) \geq 1/2$ by monotonicity of cdfs. Standard arguments on quantiles (see Corollary 21.5 in van der Vaart (1998)) yield that the median WTP has an influence function

$$r_\phi(w) = \frac{-\{1(w < \phi(\lambda_0)) - 0.5\}}{\mathbb{E}\left[\lambda_0(\phi(\lambda_0), X_i)\right]}.$$

The discontinuity of the influence function then implies that identification must be irregular, since $S^* g(w,x)$ is absolutely continuous in w. ■

Proof of Proposition 6.4. By (17), Assumption 2 holds with $\rho = 1$. By Theorem 5.1, it remains to check $I_{\theta,1} > 0$. This is, however, equivalent to $(1,0) \in \mathcal{R}(S^*)$, or existence of $g \in L_2(\mu)$ such that

$$\langle A\eta_0, g \rangle = 1, \dot{l}_\eta^* g = 0.$$

If $\langle \eta_0, g_0 \rangle \neq 0$ for g_0 one of the eigenfunctions of A^*, we define $g = c g_0$ with $c = (\theta_0 \langle \eta_0, g_0 \rangle)^{-1}$. Note that $\dot{l}_\eta^* g = c \dot{l}_\eta^* g_0 = 0$ and $\langle A\eta_0, g \rangle = c\theta_0 \langle \eta_0, g_0 \rangle = 1$. ■

Proof of Proposition 6.5. The functional $\chi(\dot{u}_0)$, although nonlinear, is concave and differentiable, with pathwise derivative

$$\dot{\chi}(b) = \mathbb{E}\left[b(C_t^*) \frac{d(C_t^*)}{\dot{u}_0(C_t^*)} \right],$$

which implies, with $\dot{l}_\eta \eta = \theta_0 A\eta(c) - \eta(c)$,

$$\begin{aligned}
\varpi(\epsilon) &= \sup_{\|\dot{u}_1 - \dot{u}_0\|_{2,\mu^*} \leq \delta, \|\dot{l}_\eta L(\dot{u}_1 - \dot{u}_0)\|_{2,\mu} \leq \epsilon} |\chi(\dot{u}_1) - \chi(\dot{u}_0)| \\
&\leq \delta \sup_{\|\dot{u}_1 - \dot{u}_0\|_{2,\mu^*} \leq \delta, \|\dot{l}_\eta L(\dot{u}_1 - \dot{u}_0)\|_{2,\mu} \leq \epsilon} |\dot{\chi}(\dot{u}_1 - \dot{u}_0)| \\
&= \delta \sup_{\|\dot{u}_1 - \dot{u}_0\|_{2,\mu^*} \leq \delta, \|\dot{l}_\eta L(\dot{u}_1 - \dot{u}_0)\|_{2,\mu} \leq \epsilon} \left| \mathbb{E}\left[\dot{l}_\eta L(\dot{u}_1(C_t) - \dot{u}_0(C_t)) g(C_t) \right] \right| \\
&\leq \delta \|g\|_{2,\mu} \epsilon,
\end{aligned}$$

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

where the first inequality uses concavity, the last equality uses (22), so

$$\begin{aligned}\dot{\chi}(\dot{u}_1-\dot{u}_0) &= \mathbb{E}\left[(\dot{u}_1-\dot{u}_0)(C_t^*)r_\chi(C_t^*)\right]\\ &= \mathbb{E}\left[(\dot{u}_1-\dot{u}_0)(C_t^*)\dot{l}_\eta^* g(C_t)\right]\\ &= \mathbb{E}\left[\dot{l}_\eta L(\dot{u}_1(C_t)-\dot{u}_0(C_t))g(C_t)\right],\end{aligned}$$

and the last inequality follows by Cauchy–Schwarz. Lemma 1.3 in the online Supplemental Appendix then implies local identification of $\chi(\dot{u}_0)$. It remains to show the regularity, but this follows from Cauchy–Schwarz, since

$$\inf_{\dot{u}:\dot{\chi}(\dot{u})\neq 0}\frac{||\dot{l}_\eta L\dot{u}||^2}{|\dot{\chi}(\dot{u})|^2} \geq \frac{\left\|\dot{l}_\eta L\dot{u}\right\|^2}{||g||_{2,\mu}^2\left\|\dot{l}_\eta L\dot{u}\right\|^2} = \frac{1}{||g||_{2,\mu}^2} > 0.$$

■

Proof of Lemma 6.1. By the change of variables with $c=\exp(z)$ and $c^*=\exp(\tau)$, and multiplying both sides by $\exp(-\tau)$, the integral equation

$$r_\chi(c^*) = \int f_\varepsilon(c/c^*)w(c)dc$$

is transformed into a convolution-type problem

$$x(\tau) = \int K(z-\tau)y(z)dz, \tag{26}$$

where $x(\tau)=\exp(-\tau)r_\chi(\exp(\tau))$, $K(u)=\exp(u)f_\varepsilon(\exp(u))$, and $y(z)=w(\exp(z))$. By Polyanin and Manzhirov (2008, p. 285), if $x\in L_2(\mathbb{R})$, a necessary and sufficient condition for existence of $y\in L_2(\mathbb{R})$ satisfying (26) is $\hat{x}/\hat{K}\in L_2(\mathbb{R})$. The solution is given by

$$y(z) = \frac{1}{2\pi}\mathrm{Re}\int e^{itz}\frac{\hat{x}(t)}{\hat{K}(t)}dt,$$

and in terms of w,

$$w(c) = \frac{1}{2\pi}\mathrm{Re}\int e^{it\log(c)}\frac{\hat{x}(t)}{\hat{K}(t)}dt.$$

Note this solution is also in $L_2(\mu)$ if $\hat{x}/\hat{K}\in L_2(\mathbb{R})$, since, by a change of variables and Fubini,

$$\begin{aligned}\mathbb{E}\left[|w(C_t)|^2\right] &= \frac{1}{(2\pi)^2}\left|\int\int\frac{\hat{x}(t)}{\hat{K}(t)}\frac{\hat{x}(s)}{\hat{K}(s)}\mathbb{E}[e^{i(t-s)\log C_t}]dtds\right|\\ &\leq \frac{1}{(2\pi)^2}\int\left|\frac{\hat{x}(t)}{\hat{K}(t)}\right|^2 dt<\infty.\end{aligned}$$

■

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit: http://dx.doi.org/10.1017/S0266466621000116

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

REFERENCES

Adams, R.A. & J.J. Fournier (2003) *Sobolev Spaces*. Academic Press.

Alan, S., O. Attanasio, & M. Browning (2009) Estimating Euler equations with noisy data: Two exact GMM estimators. *Journal of Applied Econometrics* 24, 309–324.

Altonji, J.G. & A. Siow (1987) Testing the response of consumption to income changes with noisy panel data. *Quarterly Journal of Economics* 102, 293–328.

Andrews, D.W.K. & M.A. Schafgans (1998) Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.

Bajari, P., J. Hahn, H. Hong, & G. Ridder (2011) A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models. *International Economic Review* 52, 807–824.

Begun, J.M., W.J. Hall, W.M. Huang, & J.A. Wellner (1983) Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics* 11, 432–452.

Bekker, P. & T. Wansbeek (2001) Chapter 7: Identification in parametric models. In B. Baltagi (ed.), *Companion to Theoretical Econometrics' Blackwell Companions to Contemporary Economics*, pp. 144–161. Basil Blackwell.

Bickel, P.J., C.A.J. Klassen, Y. Ritov, & J.A. Wellner (1998) *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag.

Bickel, P.J. & Y. Ritov (1988) Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya: The Indian Journal of Statistics, Series A* 50, 381–393.

Bonhomme, S. (2011) Panel Data, Inverse Problems, and the Estimation of Policy Parameters. Unpublished manuscript.

Bonhomme, S. (2012) Functional differencing. *Econometrica* 80, 1337–1385.

Carrasco, M., J.P. Florens, & E. Renault (2007) Linear inverse problem in structural econometrics estimation based on spectral decomposition and regularization. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6, pp. 5633–5751. North-Holland.

Carson, R.T. & W.M. Hanemann (2005) Contingent valuation. In K.G. Mäler & J.R. Vincent (eds.), *Handbook of Environmental Economics,* vol. 2, pp. 821–936. Elsevier.

Cattaneo, M. & J.C. Escanciano (2017) *Regression Discontinuity Designs: Theory and Applications*, Advances in Econometrics, vol. 38. Emerald Group Publishing.

Cattaneo, M.D. & M. Jansson (2019) Average Density Estimators: Efficiency and Bootstrap Consistency. Working paper, arXiv:1904.09372.

Chamberlain, G. (1986) Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics* 34, 305–334.

Chamberlain, G. (1992) Efficiency bounds for semiparametric regression. *Econometrica* 60, 567–596.

Chamberlain, G. (2010) Binary response models for panel data: Identification and information. *Econometrica* 78, 159–168.

Chen, X., V. Chernozhukov, S. Lee, & W. Newey (2014) Identification in semiparametric and nonparametric conditional moment models. *Econometrica* 82, 785–809.

Chen, X. & Z. Liao (2014) Sieve M-inference of irregular parameters. *Journal of Econometrics* 182, 70–86.

Chen, X. & S.C. Ludvigson (2009) Land of addicts? An empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics* 24, 1057–1093.

Chen, X., Pouzo, D. (2015) Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models. *Econometrica* 83, 1013–1079.

Chen, X. & M. Reiss (2011) On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* 27, 497–521.

Chen, X. & A. Santos (2018) Overidentification in regular models. *Econometrica* 86, 1771–1817.

Christensen, T.M. (2017) Nonparametric stochastic discount factor decomposition. *Econometrica* 85, 1501–1536.

Debnath, L & P. Mikusinski (2005) *Hilbert Spaces with Applications*. Elsevier Academy Press.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Dellavigna, S. & M.D. Paserman (2005) Job search and impatience. *Journal of Labor Economics* 23, 527–588.

Donoho, D.L. & R.C. Liu (1987) Geometrizing Rates of Convergence, I. Technical report 137, Department of Statistics, University of California Berkeley.

Dufour, J.M. & X. Liang (2014) Necessary and Sufficient Conditions for Nonlinear Parametric Function Identification. Unpublished manuscript.

Dynam, K. (2000) Habit formation in consumer preferences: Evidence from panel data. *American Economic Review* 90, 391–402.

Escanciano, J.C. & S. Hoderlein (2010) Nonparametric Identification of Euler Equations. Unpublished manuscript.

Escanciano, J.C., S. Hoderlein, A. Lewbel, O. Linton, & S. Srisuma (2015) Nonparametric Euler Equation Identification and Estimation. Unpublished manuscript.

Escanciano, J.C. & W. Li (2021) Optimal linear instrumental variables approximations. *Journal of Econometrics* 221, 223–246.

Fan, J. (1991) On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics* 19, 1257–1272.

Fisher, F.M. (1966), *The Identification Problem in Econometrics*. McGraw-Hill.

Giné, E. & R. Nickl (2008) A simple adaptive estimator of the integrated square of a density. *Bernoulli* 14, 47–61.

Goh, C. (2017) Rate-Optimal Estimation of the Intercept in a Semiparametric Sample-Selection Model. Working paper, arXiv:1710.01423.

Hahn, J. (1994) The efficiency bound of the mixed proportional hazard model. *Review of Economic Studies* 61, 607–629.

Hall, P. & J.S. Marron (1987) Estimation of integrated squared density derivatives. *Statistics and Probability Letters* 6, 109–15.

Hansen, L.P. & J.A. Scheinkman (2009) Long-term risk: An operator approach. *Econometrica* 77, 177–234.

Hasminskii, R.Z. & I.A. Ibragimov (1978) On the nonparametric estimation of functionals. In Proceedings of the 2nd Prague Symposium on Asymptotic Statistics, pp. 41–51. North-Holland.

Heckman, J.J. (1990) Varieties of selection bias. *The American Economic Review* 80, 313–318.

Heckman, J.J. & B. Singer (1984a) The identifiability of the proportional hazard model. *Review of Economic Studies* 51, 231–241.

Heckman, J.J. & B. Singer (1984b) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.

Hu, Y. & S.M. Schennach (2008) Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76(1), 195–216.

Hu, Y. & M. Shum (2012) Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics* 171, 32–44.

HurwiczL. (1950) Generalization of the concept of identification. In T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*. Wiley.

Ishwaran, H. (1996) Identifiability and rates of estimation for scale parameters in location mixture models. *Annals of Statistics* 24, 1560–1571.

Ishwaran, H. (1999) Information in semiparametric mixtures of exponential families. *Annals of Statistics* 27, 159–177.

Khan, S. & D. Nekipelov (2018) Information structure and statistical information in discrete response models, *Quantitative Economics*, 9(2), 995–1017.

Khan, S. & E. Tamer (2010) Irregular identification, support conditions, and inverse weight estimation. *Econometrica* 6, 2021–2042.

Koopmans, T.C. (1949) Identification problems in economic model construction. *Econometrica* 17, 125–144.

Koopmans, T.C. & O. Reirsol (1950) The identification of structural characteristics. *Annals of Mathematical Statistics* 21, 165–181.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Koševnik, Y.A. & B.Y. Levit (1976) On a non-parametric analogue of the information matrix. *Theory of Probability and Its Applications* 21, 738–753.
Kress, R. (1999) *Linear Integral Equations*. Springer.
LeCam, L. (1973) Convergence of estimates under dimensionality restrictions. *Annals of Statistics* 1, 38–53.
Lewbel, A. (1997) Semiparametric estimation of location and other discrete choice moments. *Econometric Theory* 13, 32–51.
Lewbel, A. (2019) The identification zoo—Meanings of identification in econometrics. *Journal of Economic Literature* 57(4), 835–903.
Lewbel, A., O.B.B. Linton, & D. McFadden (2011) Estimating features of a distribution from binomial data. *Journal of Econometrics* 162, 170–188.
Luenberger, D.G. (1997) *Optimization by Vector Space Methods, 1969 Edition*. Wiley.
MaCurdy, T.E. (1981) An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy* 89, 1059–1085
Manski, C. (1975) Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–228.
Matzkin, R.L. (2007) Nonparametric identification. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6b, pp. 5307–5368. Elsevier.
Matzkin, R.L. (2013) Nonparametric identification in structural economic models. *Annual Review of Economics* 5.
Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
Pfanzagl, J. (1982) *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics, vol. 13. Springer-Verlag.
Polyanin, A.D. & A.V. Manzhirov (2008) *Handbook of Integral Equations, 2nd Edition.* Chapman and Hall/CRC Press.
Pötscher, B.M. (2002) Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* 70, 1035–1065.
Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation*. Academic Press.
Ridder, G. & T. Woutersen (2003) The singularity of the efficiency bound of the mixed proportional hazard model. *Econometrica* 71, 1579–1589.
Ritov, Y. & P.J. Bickel (1990) Achieving information bounds in non and semiparametric models. *Annals of Statistics* 18, 925–938.
Rothenberg, R.J. (1971) Identification in parametric models. *Econometrica* 39(3), 577–591.
Rudin, W. (1973) *Functional Analysis*. McGraw-Hill.
Runkle, D.E. (1991) Liquidity constraints and the permanent income hypothesis. *Journal of Monetary Economics* 27, 73–98.
Samwick, A.A. (2006) Saving for retirement: Understanding the importance of heterogeneity. *Business Economics* 41, 21–27.
Santos, A. (2011) Instrumental variables methods for recovering continuous linear functionals. *Journal of Econometrics* 161, 129–146.
Sargan, J.D. (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393–415.
Sargan, J.D. (1983) Identification and lack of identification. *Econometrica* 51, 1605–1633.
Schweder, T. (1975) Window estimation of the asymptotic variance of rank estimators of location. *Scandinavian Journal of Statistics* 2, 113–126.
Severini, T.A. & G. Tripathi (2006) Some identification issues in nonparametric linear models with endogenous regressors. *Econometric Theory* 22(2), 258–278.
Severini, T.A. & G. Tripathi (2012) Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors. *Journal of Econometrics* 170(2), 491–498.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press

Shapiro, M.D. (1984) The permanent income hypothesis and the real interest rate. *Economics Letters* 14, 93–100.

Stein, C. (1956) Efficient nonparametric testing and estimation. In Proceedings of the *3rd* Berkeley Symposium on Mathematical Statistics and Probability, vol. 1. University of California Press.

van der Vaart, A.W. (1991) On differentiable functionals. *Annals of Statistics* 19, 178–204.

van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3. Cambridge University Press.

VentiS.F. & D.A. Wise (1998) The cause of wealth dispersion at retirement: Choice or chance? *American Economic Review* 88, 185–191.

https://doi.org/10.1017/S0266466621000116 Published online by Cambridge University Press