The Projection Solution to the Incidental Parameter Problem*

Andrew Chesher
UCL and CeMMAP†

Adam M. Rosen
Duke University and CeMMAP‡

Yuanqi Zhang
UCL and CeMMAP§

July 10, 2026

Abstract

This paper introduces a new approach to econometric analysis of nonlinear panel data models when the number of observations per observational unit is small. In such models the presence of variables that are constant within, while varying across, units results in an incidental parameter problem. The approach taken in this paper removes these incidental parameters *via* projection, which produces a correspondence specifying all combinations of observed variables and within-unit-varying unobserved heterogeneity that are achievable by choice of some value of the unit-specific incidental parameters. With unit-specific variables removed, there is no need for assumptions concerning their joint distribution with other variables. The result is an incomplete model which is typically partially identifying. Identified sets are characterized *via* moment inequalities using tools of random set theory. Examples of application to static and dynamic models with discrete or continuous outcomes using distribution-free restrictions on within-unit-varying unobserved heterogeneity are presented.

*We thank participants at numerous seminar and conference presentations of related work for their comments, and especially Stéphane Bonhomme, Allan Collard-Wexler, Bo Honoré, and Francesca Molinari for helpful discussion.

†Address: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Email: andrew.chesher@ucl.ac.uk.

‡Address: Adam Rosen, Department of Economics, Duke University, 213 Social Sciences Box 90097, Durham, NC 27708; Email: adam.rosen@duke.edu.

§Address: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Email: uctpyqz@ucl.ac.uk.

1 Introduction

In econometric analysis using *linear* panel data models, unit-specific heterogeneity terms, so-called "fixed effects", can be removed by differencing. This enables identification and inference without the need for restrictions on the covariation of fixed effects with observed explanatory variables and other unobserved variables, and absent parametric distributional restrictions on unobservable heterogeneity.

By contrast, in almost all *nonlinear* panel data models, differencing of observed variables or functions thereof does not remove fixed effects. This may not be an issue when observational units deliver many realizations of outcomes, enabling "Large-T" analysis. In such cases well-behaved estimators of the values of fixed effects and the values of parameters common to observational units may be available.

However, when there are few realizations per unit, as is often the case in economic data, and the values of fixed effects are estimated, estimators of common parameters can be very poorly behaved. This is the incidental parameter problem set out in Neyman and Scott (1948) and reviewed in Lancaster (2000).

With very few exceptions, approaches to this problem in nonlinear panel models restrict the joint dependence of fixed effects and other heterogeneity terms. In many empirical settings such restrictions may not be appropriate. For example, in the context of production function estimation, firm-specific heterogeneity could be related to managerial ability affecting both the level and variability of output. It may thus be desirable to allow the covariation of firm-specific and within-unit-varying heterogeneity, upon which measures of total factor productivity depend, to be unrestricted. The moment restrictions of Arellano and Bond (1991) used in the analysis of dynamic *linear* panel models have this feature, without imposing parametric restrictions on the distribution of unobservable heterogeneity. The approach in this paper enables analysis of *nonlinear* panel models that similarly leave the covariation of fixed effects with other heterogeneity unrestricted, and is applicable without imposing parametric distributional restrictions.

The approach taken here solves the incidental parameter problem by removing them from the model *via* projection. The approach is fundamentally different from others, including the functional differencing approach of Bonhomme (2012). Functional differencing finds moments that are invariant to the conditional distribution of individual effects given covariates, effectively "differencing out" the *conditional distribution* of individual effects given covariates. The functional differencing approach is only applicable in models with a parametric specification for the distribution of outcomes conditional on covariates and individual effects. The approach in this paper requires no such restrictions, because it instead projects out the individual effects themselves. With unit-specific heterogeneity removed, restrictions on its joint distribution with other variables are irrelevant.

This paper's projection approach allows great flexibility in the restrictions on unobservable heterogeneity for which identification analysis can be conducted, including nonparametric specifications of the distribution of within-unit-varying heterogeneity. The paper demonstrates with examples that feature moment conditions as in conventional GMM analysis, independence restrictions, conditional quantile restrictions, and pairwise exchangeability restrictions. The projection approach is not tied to any specific type of distributional restriction and can admit many possibilities beyond the specific cases considered here.

To explain this it helps to bring some notation on board. Consider a panel data model for outcomes $Y \equiv (Y_1, \ldots, Y_T)$ with explanatory variables $X \equiv (X_1, \ldots, X_T)$.[1] Let $U \equiv (U_1, \ldots, U_T)$ denote unobserved heterogeneity varying within units. Let V denote unit-specific unobserved heterogeneity *not* varying within units. Each element, Y_t, X_t, V and U_t can be multidimensional.[2]

Panel data models restrict the functional relationship satisfied by Y, X, V and U, defining sets of feasible values that these variables can simultaneously take. The approach proposed here works with the projection of these sets onto the space of (Y, X, U). This is the set of values of Y, X, and U that can be achieved by choice of one or more values of V. Various restrictions on the joint distribution of Y, X, and U can then be considered. With V removed, econometric analysis can proceed with no restrictions on its joint distribution with other variables.

In the linear model the projection approach delivers the set of compatible values of Y, X and U as follows.

$$\{(y, x, u) : \forall t \in \{2, \ldots, T\} \quad y_t - y_1 = (x_t - x_1)\beta + u_t - u_1\}$$

This set is defined by *equalities*, and the model is complete for differences in the outcome variables. In nonlinear models the set of compatible values of Y, X and U is typically defined by *inequalities* and there may be no nontrivial functions of outcomes for which the post-projection model is complete.

[1] It is straightforward to allow T to vary across units. In many panel applications t is an index for time, but t could index a group, family, classroom, etc.

[2] Specific to each observational-unit i there is $Y_{it}, X_{it}, V_i, U_{it}$. Observational-unit-specific indices i are omitted to simplify notation.

The paper shows how such projections can be used in identification analysis of nonlinear panel models using tools of random set theory, previously employed for identification analysis in Beresteanu, Molchanov, and Molinari (2011) and Chesher and Rosen (2017).[3] A novelty of the analysis here is the application of these tools to models that feature restrictions common in panel contexts but that are inherently absent in cross sectional settings, such as models with dynamics and models with weak exogeneity restrictions. Employing these different types of restrictions for identification analysis using random set theory is new to this paper. Identified sets for model parameters so-obtained are characterized by moment inequalities, enabling estimation and inference using approaches from the recent literature.[4]

The paper's main contribution is the projection approach to nonlinear panel models, offering a general approach for removal of incidental parameters which is not tied to any one specific kind of model (e.g. binary response) or distributional restriction. Specific examples considered here show how the analysis delivers several contributions to the nonlinear panel literature, including the following.

1. Strict and weak exogeneity restrictions allowing feedback can both be accommodated. This speaks to the emphasis in Chamberlain (2022), Bonhomme, Dano, and Graham (2023), and Bonhomme (2025) on the importance of relaxing strict exogeneity in panel models.[5] Recent developments in panel models with weak exogeneity include extension of the functional differencing approach of Bonhomme (2012) to nonlinear models with weak exogeneity in Bonhomme, Dano, and Graham (2025), and partial identification of functionals of the distribution of heterogeneous individual-specific coefficients in linear panel models in Lee (2026). The analysis in this paper contributes by allowing for weak exogeneity without placing any restrictions on the joint distribution of individual effects and t-varying heterogeneity, for example through the use of moment restrictions as in Section 3.3.

2. There is flexible treatment of initial conditions in dynamic models, distinct from the random effects treatment in Honoré and Tamer (2006). Distributional

[3] Knowledge of that theory is not required to apply the results.

[4] For example, approaches developed in Andrews and Shi (2017), Chernozhukov, Chetverikov, and Kato (2019), Bai, Santos, and Shaikh (2022), and Marcoux, Russell, and Wan (2024) can be used for asymptotic inference with uncountably many conditional moment inequalities, see also the survey Shi (2025). Characterizations based on a finite number of moment inequalities are amenable to even more approaches, see for example the recent guide Canay, Illanes, and Velez (2026).

[5] Chamberlain (2022) is a posthumously published version of a 1993 working paper.

restrictions can be conditional or unconditional on the initial condition. Unlike previous treatments, initial conditions can be unobserved since they are then unit-specific heterogeneity terms that can be removed using this paper's projection approach, as demonstrated in Section 4.1.

3. Dynamic models in which values of discrete outcomes depend upon lagged *latent* variables can be accommodated. This is useful in models in which continuous unobserved variables are coded into ordered categories as arises, for example, in studies of well-being or health status, as demonstrated in Section 4.2. Previous papers on ordered response panel models such as Honoré, Muris, and Weidner (2025) have allowed dependence on *observable* lagged outcomes, but not on lagged latent variables that determine the discrete outcomes.[6]

The paper proceeds as follows. Section 2 sets out the class of panel models studied and defines a projection of the set of feasible values of Y, X, V and U onto the space of Y, X and U. Identified sets of structures are characterized using random level sets of this projection.

Section 3 presents two examples of panel data models for continuous outcomes and derives identified sets of structural features under restrictions on the correlations amongst t-varying heterogeneity and covariates. One example involves a linear model with censored outcomes, covariates, or both. The other example concerns a CES production function in which the elasticity of substitution which appears in the production function in a nonlinear fashion is firm-specific. The moment restrictions lead to characterizations of identified sets in terms of Aumann expectations of random sets, and the use of their support functions to obtain moment inequalities.

Section 4 presents examples of dynamic panel data models for discrete outcomes and shows how to obtain identified sets for common parameters. This leads to characterizations of identified sets of structural features defined by Artstein's inequalities. There is a novel treatment of unobserved initial conditions and other missing values. A new method for accommodating autoregressive latent indexes in dynamic ordered outcome models is proposed, this by contrast to the commonly employed approaches in which outcomes follow an autoregressive process. Bounds on parameters of a dynamic binary outcome panel model are derived under quantile independence restrictions and under a conditional exchangeability restriction on the distribution of

[6]As Honoré, Muris, and Weidner (2025) note, whether it is more appropriate to model lagged dependence on the discrete outcome or on the continuous latent variable depends on the process being studied; see footnote 2 and Appendix D in that paper. We thank Bo Honoré for calling this to our attention.

U absent a parametric specification of that distribution. Section 4.5 discusses the related literature on binary and ordered panel models, and the approaches used to deal with incidental parameters in those models.

Section 5 presents numerical illustrations for the CES production function model introduced in Section 3 under both strict and weak exogeneity restrictions. Section 6 concludes.

2 The General Approach

This section first lays out the class of models covered, and then provides a general set identification characterization that will later be specialized to specific models to produce moment inequalities usable for estimation and inference.

2.1 Model, Notation, and Sampling Process

Consider a panel model specifying that

$$Y_t \in f(Y^{t-1}, X_t, V, U_t), \qquad t \in [T] \equiv \{1, ..., T\}, \tag{1}$$

for some fixed and finite T, where Y^{t-1} denotes the vector of values of Y_s for $s < t$.[7] Models allowing multi-valued functions f are accommodated by (1), thus allowing endogenous explanatory variables and models admitting multiple equilibria. The function f, all of whose arguments may be vectors, is restricted to belong to a set of functions $\mathcal{F}$, which may be parametrically or nonparametrically specified.

The panel models studied here additionally impose restrictions on the joint distribution of U and X. To incorporate such restrictions, notation $G_{U|X}(\cdot|x)$ is used to denote a conditional distribution of U conditional on $X = x$, where for any set $\mathcal{S} \subseteq \mathcal{R}_U$, $G_{U|X}(\mathcal{S}|x)$ denotes the probability of the event $U \in \mathcal{S}$ given $X = x$. Notation $G_{U|X}(\cdot|\cdot)$ denotes a collection of conditional distributions for U given $X = x$ across all possible values of $x \in \mathcal{R}_X$. Throughout the paper notation $\mathcal{R}_A$ denotes the support of any random vector A.

When considering the restrictions imposed by a model, notation $\mathsf{G}_{U|X}$ is used to denote the family of collections $G_{U|X}(\cdot|\cdot)$ admitted by the model. For instance, if the

[7]Static models are accommodated by specifying f invariant with respect to Y^{t-1}. In dynamic models initial conditions, such as Y_0 in a model with a one period lag, may either be observable, in which case these are included in Y^{t-1}, or unobservable, in which case they are included in V.

components of U are restricted to have zero mean conditional on certain components of X then $\mathsf{G}_{U|X}$ contains all such $G_{U|X}(\cdot|\cdot)$.

A sampling process delivers realizations of Y and X such that their joint distribution, F_{YX}, is identified. Realizations of V and $U \equiv (U_1, ..., U_T)$ are not observed. The former is invariant with respect to t. Variables $X \equiv (X_1, ...X_T)$ have components whose covariation with t-varying unobservable heterogeneity U is restricted. The underlying probability space on which all variables are defined is assumed nonatomic throughout.[8]

Notation $\mathcal{M}$ will be used to denote a set of pairs m of structural functions and collections of conditional distributions $m = (f, G_{U|X}(\cdot|\cdot))$ that satisfy a model's restrictions. The goal of our identification analysis is to determine which, if any, $m \in \mathcal{M}$ are capable of producing a distribution F_{YX} of observable variables.

2.2 Identification Analysis

Using the notation laid out above, the restrictions so far described are formally collected in the following restriction.

Restriction Panel Model (PM): Euclidean random vectors (Y, X, V, U) are defined on a complete nonatomic probability space $(\Omega, \mathsf{L}, \mathbb{P})$ endowed with the Borel sets on Ω such that (1) holds, where $(f, G_{U|X}(\cdot|\cdot)) \in \mathcal{M}$ and V belongs to the set $\mathcal{R}_V$. The distribution of (Y, X) is point identified. $\square$

Equivalent to (1), $Y \in \mathcal{Y}(U, V, X; f)$ almost surely, where

$$\mathcal{Y}(u, v, x; f) \equiv \left\{ y = (y_1, ..., y_T) : y_t \in f(y^{t-1}, x_t, v, u_t), \text{ all } t \in [T] \right\}. \quad (2)$$

The following defines the identified set of structures, denoted $\mathcal{I}(\mathcal{M}, F_{YX})$, delivered by a model $\mathcal{M}$ and distribution F_{YX}.

Definition 1 *Under Restriction PM the identified set of pairs of structural functions f and conditional distributions of unobservable heterogeneity $G_{U|X}(\cdot|\cdot)$ is*

$$\mathcal{I}(\mathcal{M}, F_{YX}) \equiv \Big\{ \big(f, G_{U|X}(\cdot|\cdot)\big) \in \mathcal{M} : F_{Y|X}(\cdot|x) \preceq \mathcal{Y}(\tilde{U}, \tilde{V}, X; f) \text{ for some } (\tilde{U}, \tilde{V}) \\ \text{where } \tilde{U} \sim G_{U|X}(\cdot|x) \text{ conditional on } X = x \quad a.e. \quad x \in \mathcal{R}_X \Big\}. \quad (3)$$

[8]This is a mild technical requirement on the underlying probability space that ensures convexity of the Aumann expectation of the sets $\mathcal{Q}(Y, X; \theta)$ defined in Section 3. It is assured to hold whenever U is absolutely continuously distributed with respect to Lebesgue measure. See Beresteanu, Molchanov, and Molinari (2011) for further discussion.

For any random vector A with distribution F_A and random set $\mathcal{A}$, $F_A \preceq \mathcal{A}$ denotes that F_A is selectionable with respect to the distribution of $\mathcal{A}$.[9] The definition states that the identified set comprises those $\left(f, G_{U|X}(\cdot|\cdot)\right)$ pairs in $\mathcal{M}$ for which there exist random vectors $(\tilde{U}, \tilde{V})$ with $\tilde{U}$ following the distributions of $G_{U|X}(\cdot|\cdot)$ such that the set of outcomes produced by the structural function f, namely $\mathcal{Y}(\tilde{U}, \tilde{V}, X; f)$ contains a random vector whose conditional distributions given X match those of the observed conditional distributions $\{F_{Y|X}(\cdot|x) : x \in \mathcal{R}_X\}$ almost surely. Then, and only then, the pair $\left(f, G_{U|X}(\cdot|\cdot)\right)$ are capable of producing the distribution F_{YX}.

Definition 1 follows Chesher and Rosen (2020), here incorporating both types of unobservables U and V, but it is not directly usable for estimation and inference. Making it usable requires two steps, as follows. First, it is shown that the unrestricted individual effects can be removed without loss of identifying power. Second, results from random set theory are used to yield characterizations that take the form of moment inequalities that can be used as a basis for estimation and inference.

2.2.1 Removing Individual Effects

Individual effects are removed from the model by use of the projection

$$\mathcal{R}_{YXU}(f) \equiv \left\{(y, x, u) \in \mathcal{R}_{YXU} : \exists v \in \mathcal{R}_V \text{ s.t. } y_t \in f(y^{t-1}, x_t, v, u_t), \text{ all } t \in [T]\right\}, \tag{4}$$

which is the set of values of observable variables and t-varying unobservable heterogeneity that are mutually compatible when the structural function is f.

Level sets of the projection $\mathcal{R}_{YXU}(f)$ obtained by fixing a subset of the elements (Y, X, U) are used for identification analysis. For any $(u, x, f) \in \mathcal{R}_U \times \mathcal{R}_X \times \mathcal{F}$

$$\mathcal{Y}(u, x; f) \equiv \{y \in \mathcal{R}_Y : (y, x, u) \in \mathcal{R}_{YXU}(f)\}, \tag{5}$$

is the set of possible values for y that can occur when $X = x$ and $U = u$ for some realization of V. For any $(y, x, f) \in \mathcal{R}_Y \times \mathcal{R}_X \times \mathcal{F}$

$$\mathcal{U}(y, x; f) \equiv \{u \in \mathcal{R}_U : (y, x, u) \in \mathcal{R}_{YXU}(f)\}, \tag{6}$$

is the set of possible values for u that can occur when $Y = y$ and $X = x$ for some

[9] The probability distribution of random variable A is selectionable with respect to the probability distribution of random set $\mathcal{A}$ when there exists (i) $\tilde{A}$ having the same distribution as A, and (ii) $\widetilde{\mathcal{A}}$ having the same distribution as $\mathcal{A}$, both defined on the same probability space, such that $\mathbb{P}[\tilde{A} \in \widetilde{\mathcal{A}}] = 1$. See Molchanov and Molinari (2018) Chapter 2 or Definition 2 of Chesher and Rosen (2020).

realization of V. These sets are dual to each other in that for any u, x, y, and f:

$$y \in \mathcal{Y}(u, x; f) \iff u \in \mathcal{U}(y, x; f). \tag{7}$$

There is the following Proposition.

Proposition 1 *Let Restriction PM hold. Then*

$$\mathcal{I}(\mathcal{M}, F_{YX}) = \left\{ \left(f, G_{U|X}(\cdot|\cdot)\right) \in \mathcal{M} : F_{Y|X}(\cdot|x) \preceq \mathcal{Y}(U, X; f) \text{ where } U \sim G_{U|X}(\cdot|x) \right.$$
$$\left. \text{conditional on } X = x \quad a.e. \quad x \in \mathcal{R}_X \right\}, \tag{8}$$

and equivalently,

$$\mathcal{I}(\mathcal{M}, F_{YX}) = \left\{ \left(f, G_{U|X}(\cdot|\cdot)\right) \in \mathcal{M} : G_{U|X}(\cdot|x) \preceq \mathcal{U}(Y, X; f) \text{ where } Y \sim F_{Y|X}(\cdot|x) \right.$$
$$\left. \text{conditional on } X = x \quad a.e. \quad x \in \mathcal{R}_X \right\}. \tag{9}$$

Proposition 1 provides high-level characterizations of the identified set of $\left(f, G_{U|X}(\cdot|\cdot)\right)$ from which unobservable individual effects V are absent. The random sets $\mathcal{Y}(U, X; f)$ and $\mathcal{U}(Y, X; f)$ comprise, respectively, the set of random variables Y compatible with structural function f and (U, X), and the set of random variables U possible given knowledge of observable variables (Y, X) under structural function f.[10]

When the function f is restricted to a parametric family indexed by a parameter, say θ, such that $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$, then the set-valued mappings defined in (5) and (6) will be indexed by θ rather than f, with the associated random sets denoted $\mathcal{Y}(U, X; \theta)$ and $\mathcal{U}(Y, X; \theta)$.

2.2.2 Characterization via Moment Inequalities

Proposition 1 provides high-level, generally applicable characterizations of the identified set for $\left(f, G_{U|X}(\cdot|\cdot)\right)$ in which no distributional restrictions are placed on unobservable individual effects V.[11] To make these characterizations practicable requires necessary and sufficient conditions for the stated selectionability properties usable

[10] The selectionability statement requiring $G_{U|X}(\cdot|x)$ to be selectionable with respect to the distribution of $\mathcal{U}(Y, X; f)$ conditional on X almost surely in (9) in Proposition 1 is equivalent to requiring that (U, X) is selectionable with respect to the distribution of the random set $\mathcal{U}(Y, X; f) \times \{X\}$ by Proposition 1 in Appendix B of Chesher and Rosen (2015). An analagous statement holds regarding selectionability of $F_{Y|X}(\cdot|x)$ with respect to the distribution of $\mathcal{Y}(U, X; f)$ almost surely in (8), and selectionability of (Y, X) with respect to the distribution of $\mathcal{Y}(U, X; f) \times \{X\}$.

[11] Characterization of the identified set for f or any functional of $\left(f, G_{U|X}(\cdot|\cdot)\right)$ then follows.

for estimation and inference. The recent literature employing random set theory for identification analysis has made use of such conditions, which can be employed here as well.[12] One approach uses Artstein's inequality under the following mild restriction.

Restriction RCS: For all $\left(f, G_{U|X}(\cdot|\cdot)\right) \in \mathcal{M}$, $G_{U|X}\left(\mathsf{cl}\left(\mathcal{U}(y,x;f)\right) \setminus \mathcal{U}(y,x;f)|x\right) = 0$ *a.e.* (y,x), where $\mathsf{cl}(\cdot)$ and $\cdot \setminus \cdot$ denote the closure of a set, and the set difference of two sets, respectively. $\square$

Restriction RCS holds automatically if $\mathcal{U}(Y,X;f)$ is closed almost surely. It also applies when that is not so, but the difference between $\mathcal{U}(Y,X;f)$ and its closure is measure zero almost surely.[13] In this case the restriction enables characterization of identified sets by applying results from random set theory to the closure of $\mathcal{U}(Y,X;f)$, which is useful since selectionability criteria from random set theory are often stated for random closed sets. Lemma 1 in the Appendix provides the formal statement.

2.2.3 Additional Notation

For any random vector A or realized value a let $A_{st}^{\Delta} \equiv A_s - A_t$ and $a_{st}^{\Delta} \equiv a_s - a_t$. For any integer $k > 0$, notation $\mathbf{0}_k$ denotes a zero vector of length k. In the definition of any set, such as $\mathcal{U}(y,x;f)$ in (6) above, the support of a variable is omitted when it is clear from context. Notation expressing suprema and infima of conditional probabilities or expectations with respect to the conditioning variables are to be understood as *essential* suprema and infima, respectively. For any real number c, $c^- \equiv |\min\{c,0\}|$ and $c^+ \equiv \max\{c,0\}$ denote the negative and positive part of c, respectively. For random vectors A and B, $A \perp\!\!\!\perp B$ signifies that A and B are stochastically independent. For any vectors a and b, $a \cdot b$ denotes their dot product.

3 Models with continuous outcomes

This section considers two models with essential nonlinearity and continuous outcomes. It is shown how the projection approach leads to identification results and so to estimation and inference absent restrictions involving the distribution of unit-specific variables and absent a parametric specification of the distribution of within-unit-varying unobservables. Identification results are obtained under moment restrictions

[12] For further details and alternative conditions for guaranteeing this *selectionability* property see Chapter 2 of Molchanov and Molinari (2018).

[13] This is a common occurrence in models in which inequalities determine the value of a limited dependent variable, for example when a discrete outcome is determined by whether a continuously distributed unobservable variable exceeds a threshold.

involving within-unit-varying unobservables and functions of covariates.

In Sections 3.1 and 3.2 the results of projection are derived. In Section 3.3 identification analysis using these projections is provided.

3.1 CES Production Function Panel Models

This is an example of a nonlinear panel model with continuous outcomes, inspired by Example 3 of Bonhomme (2012). Let log output Y_t of an individual unit (e.g. a plant or firm) at time t be generated by a constant elasticity of substitution (CES) production function such that

$$Y_t = \beta \log g(L_t, K_t, \gamma, S) + C + U_t, \qquad t \in [T], \tag{10}$$

where $V = (S, C)$ is a pair of unit-specific unobservable variables, $C \in \mathbb{R}$, and

$$g(l, k, \gamma, s) \equiv \left(\gamma l^s + (1-\gamma)k^s\right)^{1/s}, \quad l, k > 0, \quad s \in [-\infty, 1], \quad \gamma \in (0, 1)\,,$$

is the CES function with substitution parameter s.[14]

Variables $X_t \equiv (L_t, K_t)$ denote labor and capital inputs at time t, U_t denotes t-varying unobservable variables, and $\theta = (\beta, \gamma)$ are common parameters with $\beta > 0$.

Thus (1) holds for $f = f_\theta$ with

$$f_\theta(Y^{t-1}, X_t, V, U_t) = \beta \log g(L_t, K_t, \gamma, S) + C + U_t. \tag{11}$$

This specification is considered briefly in Bonhomme (2012) as an example of nonlinear regression, with a unit-specific effect (S) entering nonlinearly.[15] In that paper $U \equiv (U_1, \ldots, U_T)$ is restricted Gaussian independent of (X, V), where V is denoted by α. It is stated on page 1344, "Due to the nonlinearity, it does not seem possible to difference out α in a straightforward way."

The projection approach can be applied here, as is now illustrated. The Gaussian restriction proposed in Bonhomme (2012) is not required.

The set of values of (Y, X, U) delivered by the model for some value of V is obtained as follows.

The function $g(l, k, \gamma, s)$ is a generalized mean, monotone increasing in $s \in [-\infty, 1]$,

[14] So $\mathcal{R}_S$ is the subset of the extended real line on which all elements are no greater than one.

[15] Equation (10) appears under Example 3 as equation (6) in Bonhomme (2012), in the notation of that paper using the symbol σ where we use S. The discussion here uses a simplified version in which there is no low-skilled labor input.

bounded with

$$g(l,k,\gamma,-\infty) \equiv \min\{l,k\}, \qquad g(l,k,\gamma,0) \equiv l^{\gamma} k^{1-\gamma}.$$

With $x_t \equiv (x_{t1}, x_{t2}) = (l_t, k_t)$ and $v = (s, c)$ define for each t:

$$m_t(y,x,\theta,v) \equiv y_t - \beta \log g(x_{t1}, x_{t2}, \gamma, s) - c, \tag{12}$$

which is monotone decreasing in each component of v.

The set of feasible combinations of (y, x, u) obtained by projection across v is

$$\mathcal{R}_{YXU}(\theta) = \big\{(y,x,u) : \exists v \in [-\infty, 1] \times \mathbb{R} \text{ s.t. } u_t = m_t(y,x,\theta,v) \text{ all } t \in [T]\big\}. \tag{13}$$

The U-level set of this projection for any realizations y, x is

$$\mathcal{U}(y,x,\theta) = \big\{u : \exists v \in [-\infty, 1] \times \mathbb{R} \text{ s.t. } u_t = m_t(y,x,\theta,v) \text{ all } t \in [T]\big\}. \tag{14}$$

If V were observable then realization of $V = v$ would reveal the realization of U as the singleton value with components $m_t(y,x,\theta,v)$, for each $t \in [T]$. The set $\mathcal{U}(y,x,\theta)$ is a manifold on $\mathbb{R}^T$ comprising the set of such vectors compatible with *some* value of unobservable V.

For ease of illustration consider the case in which $T = 3$. Manifolds $\mathcal{U}(y,x;\theta)$ are illustrated for an example in which $x = (l_t, k_t)$, $t = 1, 2, 3$ are set according to:

$$l_1 = 0.5, \quad l_2 = 1, \quad l_3 = 2, \quad k_1 = 2.2, \quad k_2 = 1.2, \quad k_3 = 0.7,$$

and parameters are set at $\gamma = 0.6$ and $\beta = 1$. The left panel of Figure 1a depicts eight manifolds $\mathcal{U}(y,x;\theta)$, one each for values of y with y_{21}^{Δ} and y_{31}^{Δ} taking values shown in green in the right panel.[16] Figure 1b shows the set $A(\theta)$ in blue in the left hand panel comprising the union of the sets $\mathcal{U}(y,x;\theta)$ obtained as y takes values such that y_{21}^{Δ} and y_{31}^{Δ} belong to the region B shaded in green in the right hand panel. Because $(Y_{21}^{\Delta}, Y_{31}^{\Delta}) \in B$ implies that $(U_{21}^{\Delta}, U_{31}^{\Delta}) \in A(\theta)$, there is for all $x \in \mathcal{R}_X$ the inequality

$$\mathbb{P}[(Y_{21}^{\Delta}, Y_{31}^{\Delta}) \in B | X = x] \leq \mathbb{P}[(U_{21}^{\Delta}, U_{31}^{\Delta}) \in A(\theta) | X = x],$$

which restricts the values of θ compatible with F_{YX} because of the dependence of

[16] Because the individual effect C enters additively, each manifold $\mathcal{U}(y,x,\theta)$ is represented in the space of differences $(u_{21}^{\Delta}, u_{31}^{\Delta})$ and is fully determined by the values of y_{21}^{Δ} and y_{31}^{Δ}.

$A(\theta)$ on θ. Implications of this sort also arise by use of Artstein's inequality. If U and X are stochastically independent the inequality above becomes

$$\sup_{x \in \mathcal{R}_X} \mathbb{P}[(Y_{21}^{\Delta}, Y_{31}^{\Delta}) \in B | X = x] \leq \mathbb{P}[(U_{21}^{\Delta}, U_{31}^{\Delta}) \in A(\theta)].$$

Section 3.3 characterizes identified sets for the common parameters obtained in this model under moment restrictions on the product of elements of U and functions or components of X. First, the following section presents a second example of a panel model with continuous outcomes.

3.2 Linear Panels with Censored Outcomes and Covariates

This section provides results for linear panel models when data are interval censored.[17]

The model specifies

$$Y_t^* = X_t^* \theta + C + U_t, \qquad t \in [T], \tag{15}$$

where θ is a $k_x \times 1$ vector of common parameters, each X_t^* is a $1 \times k_x$ vector, and $V = C$. The unobserved variables are Y^*, X^*, C, and U. The observed variables are

$$X \equiv \{(X_t^L, X_t^H) : t \in [T]\}, \quad Y \equiv \{\left(Y_t^L, Y_t^H\right) : t \in [T]\}, \tag{16}$$

where

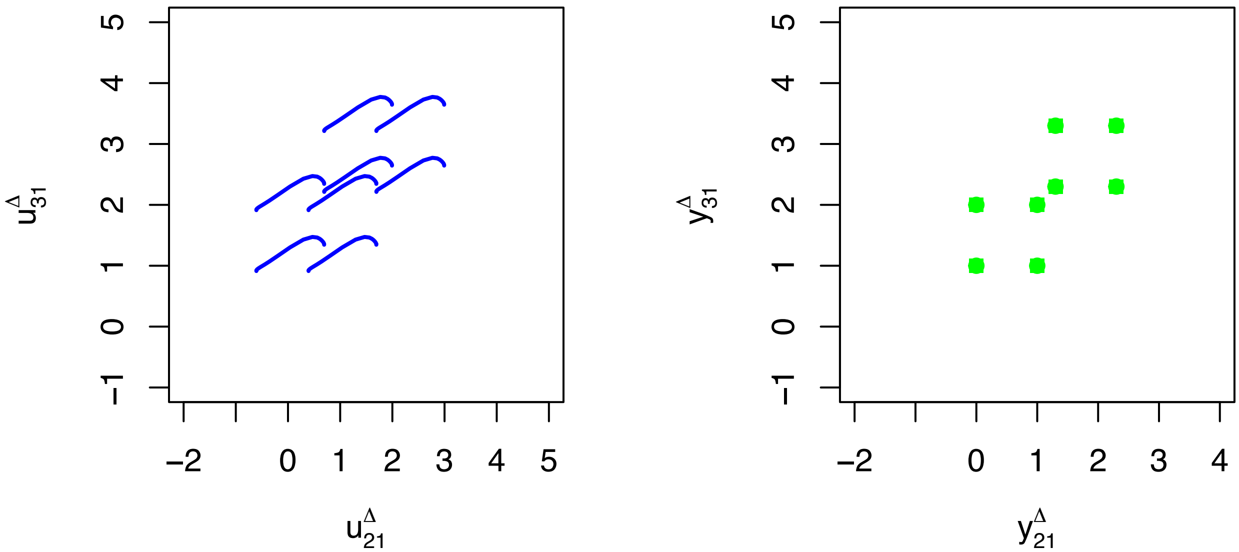
$$\forall t \in [T]: \qquad Y_t^L \leq Y_t^* \leq Y_t^H, \qquad X_t^L \leq X_t^* \leq X_t^H, \tag{17}$$

and inequalities hold element-wise. There may be interval censoring of components of one or both of the outcome Y^* and covariates X^*. Components that are not censored have $Y_t^L = Y_t^H$ for outcomes and $X_{tj}^L = X_{tj}^H$ for any uncensored components X_{tj}^* of X_t^*. Missing data can be captured by having both lower and upper limits correspond to the end points of the support of the corresponding variables.

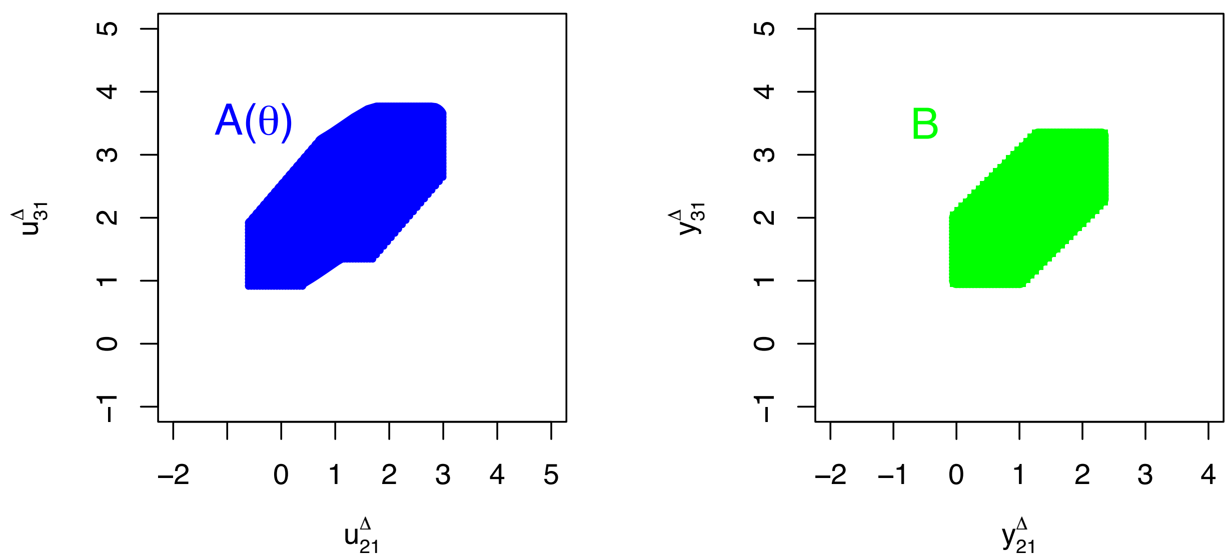
Define

$$B_t^L(Y, X; \theta) \equiv Y_t^L - \sum_{j=1}^{k_x} \max\{X_{tj}^H \theta_j, X_{tj}^L \theta_j\},$$
$$B_t^H(Y, X; \theta) \equiv Y_t^H - \sum_{j=1}^{k_x} \min\{X_{tj}^H \theta_j, X_{tj}^L \theta_j\},$$

[17] Analysis of panel models with censored outcomes has been studied in e.g. Honoré (1992, 1993), Hu (2002), Khan, Ponomareva, and Tamer (2016), and Abrevaya and Muris (2020).



(a) The panel above on the left illustrates eight manifolds $\mathcal{U}(y, x; \theta)$ for the CES production function example with x and θ as described in the text. The panel above on the right shows the corresponding values of y_{21}^{Δ} and y_{31}^{Δ}.



(b) The panel above on the left illustrates the union of the sets $\mathcal{U}(y, x; \theta)$ as y varies across all values of y such that y_{21}^{Δ} and y_{31}^{Δ} take values in the set shown on the right for the CES production function example with x and θ as described in the text.

Figure 1: CES production function manifolds $\mathcal{U}(y, x; \theta)$.

There is, for all s and t in $[T]$:

$$\begin{aligned} B_t^L(Y,X;\theta) &\leq C + U_t \quad \leq B_t^H(Y,X;\theta), \\ -B_s^H(Y,X;\theta) &\leq -C - U_s \leq -B_s^L(Y,X;\theta). \end{aligned}$$

Adding the two inequalities yields the projection of the model-admitted set of values of (Y, X, C, U) onto the space of (Y, X, U) as follows.

$$\mathcal{R}_{YXU}(\theta) = \left\{(y,x,u) : \forall s,t, \quad B_t^L(y,x;\theta) - B_s^H(y,x;\theta) \leq u_t - u_s \leq B_t^H(y,x;\theta) - B_s^L(y,x;\theta)\right\} \tag{18}$$

Recall that, absent censoring, the linear model by contrast delivers the projected set

$$\mathcal{R}_{YXU}(\theta) = \{(y,x,u) : \forall s,t, \quad u_t - u_s = y_t - y_s - (x_t - x_s)\theta\} \tag{19}$$

in which there are equalities, whereas with censoring there are inequalities as in the CES production function case.

With censoring the level set of U-values that deliver $Y = y$ when $X = x$ is simply the slice through the projection defined in (18) obtained fixing (y, x) accordingly:

$$\mathcal{U}(y,x;\theta) = \{u : \forall s,t, \quad B_t^L(y,x;\theta) - B_s^H(y,x;\theta) \leq u_t - u_s \leq B_t^H(y,x;\theta) - B_s^L(y,x;\theta)\}.$$

This characterization of $\mathcal{U}(Y, X;\theta)$ provides a starting point for identification analysis using various restrictions on the joint distribution of (X, U). When covariates are censored, consideration of context may lead one to prefer restrictions on the joint distribution of (X^*, U), as considered in the following subsection.

3.3 Identified sets

Identification analysis for common parameters is now presented for the two models just considered under moment restrictions on within-unit-varying unobservables and covariates. A characterization for the CES model parameters under a conditional mean restriction is then provided. The methods employed can be applied to more general forms of moment conditions than the ones considered here.

3.3.1 Moment Restrictions

Consider the following restriction.

Restriction M. For all $t \in [T]$, $E[Z_t U_t] = \mathbf{0}_{J_t}$, where each $Z_t = Z_t(X)$ is a vector-valued function of X of dimension J_t.

Weak exogeneity and strict exogeneity of components of X can be accommodated by appropriate definition of each Z_t as will be shown below. Flexibly defining Z_t can further be used to specify that some covariates are strictly exogenous while others are only weakly exogenous.

Since Restriction M requires that $E[Z_t U_t] = \mathbf{0}_{J_t}$ for all t, it is useful to define the level set of possible values of $((Z_1 U_1), ..., (Z_T U_T))$ obtained from the projection $\mathcal{R}_{YXU}(\theta)$ defined in (4), making use of the U-level sets defined in (6):

$$\mathcal{Q}(y, x; f) \equiv \left\{ \left((z_1 u_1), ..., (z_T u_T) \right) : u \in \mathcal{U}(y, x; f) \right\}. \tag{20}$$

Thus $\mathcal{Q}(y, x; f)$ is a set of vectors, each element of which is a $J \equiv J_1 + \cdots + J_T$ dimensional vector whose entries correspond to feasible values of components of $Z_t U_t$ under structural function f across all $t \in [T]$, when $Y = y$ and $X = x$.

Replacing fixed arguments (y, x) with (Y, X) in (20) yields $\mathcal{Q}(Y, X; f)$, a random set whose distribution is determined by that of (Y, X). Under Restriction M, the panel model specification (1) with structural function f can produce the distribution of (Y, X) if and only if there is a *measurable selection* of $\mathcal{Q}(Y, X; f)$, defined below, whose expected value is the zero vector $\mathbf{0}_J$.

Definition 2 *Let $\mathcal{Q}$ be a random closed set on $(\Omega, \mathsf{L}, \mathbb{P})$ whose realizations are subsets of $\mathbb{R}^J$. A random vector Q measurable on $(\Omega, \mathsf{L}, \mathbb{P})$ is a **measurable selection** of $\mathcal{Q}$ if $Q(\omega) \in \mathcal{Q}(\omega)$ for almost all $\omega \in \Omega$.*

Let $\mathbf{L}^1(\mathcal{Q})$ denote the set of all integrable measurable selections of a random set $\mathcal{Q}$. The Aumann integral and Aumann expectation of $\mathcal{Q}$ are defined, respectively, as

$$\mathbb{E}_I[\mathcal{Q}] \equiv \left\{ E[Q] : Q \in \mathbf{L}^1(\mathcal{Q}) \right\}, \qquad \mathbb{E}[\mathcal{Q}] \equiv \mathsf{cl}\left(\mathbb{E}_I[\mathcal{Q}]\right).$$

Define the sets

$$\mathcal{F}_I \equiv \{ f \in \mathcal{F} : \mathbf{0}_J \in \mathbb{E}_I [\mathcal{Q}(Y, X; f)] \}, \quad \overline{\mathcal{F}_I} \equiv \{ f \in \mathcal{F} : \mathbf{0}_J \in \mathbb{E}[\mathcal{Q}(Y, X; f)] \}. \tag{21}$$

Under the restrictions of Proposition 2, the set $\mathcal{F}_I$ is the identified set for structural function f. The set $\overline{\mathcal{F}_I}$ is a superset of $\mathcal{F}_I$, and therefore provides bounds on f.

The set $\overline{\mathcal{F}_I}$ is the *moment-closure* of the identified set in the terminology of Li (2026), and it is useful for estimation and inference for several reasons. First, the set

$\overline{\mathcal{F}_I}$ can coincide with $\mathcal{F}_I$ and hence be sharp, for example when the Aumann integral is closed, in which case the Aumann integral and Aumann expectation coincide. Such conditions hold under a mild restriction for the CES example of Section 3.1, as shown in Proposition 3 in the Appendix. Second, characterization of $\overline{\mathcal{F}_I}$ by way of the Aumann expectation is equivalent to a support function characterization which has the form of moment inequalities that only involve expectations of random variables rather than random sets. Third, there are conditions under which the identified set and its moment closure are statistically indistinguishable even if they do not coincide, as shown in Li (2026).

The usefulness of the support function characterization for the moment closure follows from two consequences of the restrictions of Proposition 2 below. First, there is the equivalence

$$\mathbf{0}_J \in \mathbb{E}\left[\mathcal{Q}(Y,X;f)\right] \iff \min_{r\in\mathcal{B}^J} h\left(\mathbb{E}\left[\mathcal{Q}(Y,X;f)\right], r\right) \geq 0, \tag{22}$$

where $\mathcal{B}^J \equiv \left\{r \in \mathbb{R}^J : \|r\| = 1\right\}$ is the boundary of the unit ball in $\mathbb{R}^J$ and

$$h(\mathcal{Q}, r) \equiv \sup_{q\in\mathcal{Q}} r \cdot q,$$

denotes the support function of any set $\mathcal{Q} \subseteq \mathbb{R}^J$ evaluated at $r \in \mathbb{R}^J$.[18] Second, the order of the expectation and support function can be swapped yielding

$$h\left(\mathbb{E}\left[\mathcal{Q}(Y,X;f)\right], r\right) = E\left[h\left(\mathcal{Q}(Y,X;f), r\right)\right],$$

where on the right there is the expectation of a random variable.[19] Putting all this together, there is the following Proposition regarding $\mathcal{F}_I$ and $\overline{\mathcal{F}_I}$ defined in (21).

Proposition 2 *Suppose that Restrictions PM and M hold and that $\mathcal{Q}(Y,X;f)$ is closed almost surely. Then $\mathcal{F}_I$ is the identified set for the structural function f. Moreover, the moment closure $\overline{\mathcal{F}_I}$ of $\mathcal{F}_I$ comprises bounds on f and admits the support function representation*

$$\overline{\mathcal{F}_I} = \left\{f \in \mathcal{F} : \min_{r\in\mathcal{B}^J} E\left[h(\mathcal{Q}(Y,X;f), r)\right] \geq 0\right\}. \tag{23}$$

[18]This follows from Theorem 2.1.26 of Molchanov (2017) because the underlying probability space is nonatomic under Restriction PM and $\mathcal{Q}(Y,X;\theta)$ is integrable under Restriction M, so $\mathbb{E}\left[\mathcal{Q}(Y,X;\theta)\right]$ is convex.

[19]This follows from Theorem 2.1.35 of Molchanov (2017) because the underlying probability space is nonatomic.

If $\mathbb{E}_I\left[\mathcal{Q}(Y,X;f)\right]$ *is closed then* $\overline{\mathcal{F}_I}=\mathcal{F}_I$.

Proposition 2 states that $\mathcal{F}_I$ is the identified set for f, and that its moment closure $\overline{\mathcal{F}_I}$ is characterized by the moment inequalities $E\left[h(\mathcal{Q}(Y,X;f),r)\right]\geq 0$, for all $r\in\mathbb{R}^J$. The reasoning is similar to that of Theorem 4.1 of Beresteanu, Molchanov, and Molinari (2011) which established sharp bounds on the best linear predictor with censored outcomes and covariates. The difference is that the analysis here applies to a nonlinear panel model with individual effects, rather than a model for cross section data, so the random set $\mathcal{Q}(Y,X;f)$ is constructed by taking the product of components of Z as specified by Restriction M with the U-level set obtained from $\mathcal{R}_{YXU}(f)$ after removing V from the model by projection.[20]

A connection to the support function approach of Beresteanu, Molchanov, and Molinari (2011) is also made in Lee (2026) which, in contrast to the projection approach, uses moments that restrict the joint distribution of individual effects with other heterogeneity terms. That paper uses duality theory for infinite dimensional programs to characterize the identified set of features of the distribution of random coefficients in linear panel models. It also shows that if the target parameter is a common parameter, the characterization is equivalent to that obtained by the support function approach used in Beresteanu, Molchanov, and Molinari (2011) Theorem 4.1. Thus, subject to regularity conditions, duality theory for infinite dimensional programs can also be used with the moment conditions of this paper for identification analysis. This relationship also highlights the possibility of using the formulation of Schennach (2014) for developing estimation and inference approaches as a potential alternative to the moment inequality approach, an avenue which is left to future research.

Identification in the CES Model

In the CES model the class $\mathcal{F}$ is parameterized by $\theta=(\beta,\gamma)$. Focus is given to the moment closure of the identified set for θ, denoted Θ^*, and the resulting moment inequalities. The notation replaces f with θ accordingly. In the CES model

$$\mathcal{Q}(y,x;\theta)\equiv\left\{\left(z_1m_1\left(y,x;\theta,v\right)\right),...,\left(z_Tm_T(y,x;\theta,v)\right)\right):v\in[-\infty,1]\times\mathbb{R}\right\},\qquad(24)$$

[20]The projection step renders an additional integrably boundedness condition used in Theorem 4.1 of Beresteanu, Molchanov, and Molinari (2011) inapplicable without further restrictions in the present setting, necessitating the distinction between the identified set and its moment closure.

where $m_t(y, x, \theta, v)$ defined in (12) denotes for each t the unique value of u_t given fixed values of (y, x, θ, v).

The characterization of Proposition 2 is specialized under Restriction M with two different specifications for Z as follows.

Restriction MS. Restriction M holds with $Z_t = (1, X_1, ..., X_T)$ for all t.

Restriction MW. Restriction M holds with $Z_t = (1, X_1, ..., X_t)$ for all t.

Consider first Restriction MS which requires that $E[X_{hk}U_t] = 0$ for all $h, t \in [T]$ and all $k = 1, 2$ where $X_{h1} = L_h$ and $X_{h2} = K_h$. This is a strict exogeneity restriction comprising $J = T(2T + 1)$ moment restrictions. The expectation of the support function of $\mathcal{Q}(Y, X; \theta)$ given by (24) under the CES specification simplifies as

$$E[h(\mathcal{Q}(Y, X; \theta), r)] = E\left[\sup_v \sum_{t=1}^{T} w_t(r, X) m_t(Y, X, v, \theta)\right],$$

where for each $t \in [T]$,

$$w_t(r, x) \equiv r_t + \sum_{h=1}^{T} \sum_{k=1}^{2} r_{thk} x_{hk},$$

and r is a list of $J = T(2T + 1)$ numbers r_t, r_{thk} for all t, h, k.

Using the definition of m_t in (12) and simplifying, it follows from Proposition 2 that under Restriction MS the set Θ^* is the set of $\theta = (\beta, \gamma)$ that satisfy

$$\min_{r:\|r\|=1} E\left[\sup_{s,c} \sum_{t=1}^{T} w_t(r, X) \left(Y_t - \beta \log g(X_{t1}, X_{t2}, \gamma, s) - c\right)\right] \geq 0. \tag{25}$$

Now suppose instead that only weak exogeneity is asserted, such that Restriction MW is imposed. Working through the same steps under this weaker restriction, Proposition 2 delivers Θ^* as those parameter vectors satisfying the same inequalities (25), but now with r additionally restricted to satisfy $r_{thk} = 0$ for all $t < h$. Minimization over this restricted set imposes the zero moment restrictions $E[X_{hk}U_t] = 0$ only for $t \geq h$.

Section 5 demonstrates that Θ^* can produce informative sets by way of numerical illustrations under both weak and strict exogeneity restrictions. In that section further simplification of the inequalities (25) is provided for that purpose.

Moment restrictions on other functions of covariates and components of U may similarly be imposed through Restriction M beyond the two specific cases of Restriction MS and Restriction MW considered here. Moment conditions incorporating instrumental variables can be used by specifying components of Z_t as functions of components of X with respect to which the structural function is restricted to be invariant, for example by way of exclusion restrictions.

Identification in the Censored Linear Panel Model

Moment restrictions can also be used in the censored linear panel model described in Section 3.2. Focus is again given to the moment closure of the identified set for θ, denoted Θ^*, and the resulting moment inequalities. In this model it may be desirable to invoke moment restrictions involving functions of censored covariate values X^* rather than the observed endpoints of intervals on which X^* is realized, which allows censoring to be endogenous.[21] Thus the following restriction is considered.

Restriction M*: For all $t \in [T]$, $E\left[Z_t U_t\right] = \mathbf{0}_{J_t}$, where each $Z_t = Z_t(X^*)$ is a vector-valued function of X^* of dimension J_t.

The level set $\mathcal{Q}(y, x; \theta)$ of possible values of $\left((Z_1 U_1), ..., (Z_T U_T)\right)$ obtained from the censored linear panel projection $\mathcal{R}_{YXU}(\theta)$ given in (19) under Restriction M^* is

$$\left\{\left(Z_1(x^*)(y_1^* - x_1^*\theta - c), ..., Z_T(x^*)(y_T^* - x_T^*\theta - c)\right) : c \in \mathcal{R}_C, (y^*, x^*) \in \mathcal{D}(y, x)\right\},$$

where $\mathcal{D}(y, x)$ denotes the set of possible values of the censored variables given (y, x):

$$\mathcal{D}(y, x) \equiv \left\{(y^*, x^*) : y_t^L \leq y_t^* \leq y_t^H, \quad x_t^L \leq x_t^* \leq x_t^H, \text{ all } t \in [T]\right\}.$$

Following the same reasoning used when considering Restriction M and the CES model the moment closure of the identified set, Θ^*, comprises θ such that $\mathbf{0}_J \in \mathbb{E}\left[\mathcal{Q}(Y, X; \theta)\right]$. Use of the support function yields an equivalent characterization *via* moment inequalities:

$$\min_{r:\|r\|=1} E\left[\sup\left\{\sum_{t=1}^{T}\left(\overrightarrow{r_t} \cdot \left(Z_t(x^*)(y_t^* - x_t^*\theta - c)\right)\right) : c \in \mathcal{R}_C,\ (y^*, x^*) \in \mathcal{D}(y, x)\right\}\right] \geq 0,$$

where each $\overrightarrow{r_t}$ is a vector of length J_t such that $r = (\overrightarrow{r_1}, ..., \overrightarrow{r_T})$.

As was the case for Restriction M, Restriction M* can accommodate different moment restrictions through specification of Z_t. For example, $Z_t = (1, X_1^*, ..., X_T^*)$ and $Z_t = (1, X_1^*, ..., X_t^*)$ for strict and weak exogeneity restrictions, respectively.

[21] If moment restrictions are made solely with respect to X, analysis following the steps of the previous section incorporating the linear panel specification applies directly, so is not repeated here.

3.3.2 Conditional Moment Restrictions

Consider the following conditional moment restriction, which provides a strict exogeneity restriction stronger than that of Restriction MS.

Restriction CMS: For all $t \in [T]$, $E[U_t|X] = 0$.

Restriction CMS implies that $\mathbf{0}_T \in \mathbb{E}[\mathcal{U}(Y,X;\theta)|X]$ almost surely, i.e. that the zero vector is an element of the conditional Aumann Expectation of the U-level set, the form of which for the CES model is given in (14).

Using the support function approach in the CES model this is equivalently that

$$\min_{\|r\|=1} E\left[\sup_{s\in[-\infty,1]} \sup_{c\in\mathbb{R}} \sum_{t=1}^{T} r_t\big(Y_t - \beta \log g(X_{t1}, X_{t2}, \gamma, s) - c\big)|X = x\right] \geq 0, \text{ a.e. } x,$$

where $r = (r_1, ..., r_T)$. This can further be expressed

$$\min_{\|r\|=1} \left(\sum_{t=1}^{T} r_t \mu_t(x) - \min_{s\in[-\infty,1]} \sum_{t=1}^{T} r_t \beta \log g(x_{t1}, x_{t2}, \gamma, s) - \inf_{c\in\mathbb{R}} c \sum_{t=1}^{T} r_t\right) \geq 0, \text{ a.e. } x,$$

where $\mu_t(x) \equiv E[Y_t|X]$. From this it follows that for any θ only values of $r_1, ..., r_T$ that sum to zero can provide the minimum over r, and the inequalities simplify to

$$\min_{\{r:\|r\|=1, r_1+\cdots+r_T=0\}} \left(\sum_{t=1}^{T} r_t \mu_t(x) - \min_{s\in[-\infty,1]} \sum_{t=1}^{T} r_t \beta \log g(x_{t1}, x_{t2}, \gamma, s)\right) \geq 0, \text{ a.e. } x,$$

So, as in the cases considered with unconditional moment restrictions, the bounds are characterized by an infinite collection of moment inequalities, in this case infinitely many conditional moment inequalities. As noted in the introduction estimation and inference methods from the recent literature are available, as discussed for example in the survey Shi (2025).

4 Models with discrete outcomes

This section gives examples of application of the projection approach in static and dynamic panel models with *discrete* outcomes. A dynamic binary outcome model is studied in Section 4.1; a dynamic ordered response model is studied in Section 4.2.

It is shown how, in models in which ordered discrete outcomes encode the values of continuous latent variables, autoregressive dependence in the latent continuous variables can be accommodated. This approach can also be employed in other discrete response models, such as multinomial choice.

In dynamic models attention must be paid to initial values of outcomes if they are not observed. Following the approach of this paper, when they are not observed they are treated as unit-specific unobserved variables and are removed by projection. This is illustrated in the models considered in both Sections 4.1 and 4.2.

Section 4.3 provides characterizations of identified sets for the binary and ordered response models of Sections 4.1 and 4.2. Moment-based restrictions such as those considered in continuous outcome models in Section 3 can be uninformative in discrete outcome models; see Manski (1988). So here attention is turned to the identifying power of stochastic independence restrictions. Section 4.4 considers further options for conducting identification analysis absent a parametric specification of the distribution of t-varying unobservables, such as quantile and exchangeability restrictions.

Section 4.5 provides discussion of the related literature on binary and ordered outcome panel models. In contrast to the approach here, nearly all such models in the literature restrict the joint distribution of unit-specific effects and within-unit-varying heterogeneity.[22] To our knowledge, no previous models allow for unobserved initial conditions or for dependence on lagged *latent* variables.

4.1 A Dynamic Binary Outcome Panel Model

Consider the following binary outcome panel specification.

$$Y_t = 1\left[X_t\beta + \gamma Y_{t-1} + C + U_t \geq 0\right], \qquad t \in [T], \tag{26}$$

with parameter vector $\theta = (\gamma, \beta')$, $\mathcal{R}_C = \mathbb{R}$, and U continuously distributed with full support on $\mathbb{R}^T$ conditional on X.[23] Higher order lags are easily accommodated.

In a dynamic model in which γ may be nonzero, there is an initial condition to be considered. If the value of Y_0 is observable, then the t-invariant individual unobservable is $V = C$, while if Y_0 is not observable $V = (C, Y_0)$. Whichever is the case, define $\mathcal{Y}_0$ to be the set of possible values of the initial condition such that, if the initial condition is observable, then $\mathcal{Y}_0 = \{y_0\}$ and if it is not then $\mathcal{Y}_0 = \{0, 1\}$.

[22] The analysis in Aristodemou (2021) is the sole exception of which we are aware.

[23] This implies that Restriction RCS holds and there is no loss in using weak inequalities throughout in expressions for the sets $\mathcal{R}_{YXU}(\theta)$ and $\mathcal{U}(y, x; \theta)$.

Projecting away individual effects yields

$$\mathcal{R}_{YXU}(\theta) = \{(y, x, u) : \exists (y_0, c) \in \mathcal{Y}_0 \times \mathbb{R} : y_t = 1\left[x_t\beta + \gamma y_{t-1} + c + u_t \geq 0\right],\ t \in [T]\}.$$

It is convenient to define sets of indices

$$\mathcal{T}_0 \equiv \{t \in [T] : Y_t = 0\}, \qquad \mathcal{T}_1 \equiv \{t \in [T] : Y_t = 1\}.$$

Observing that

$$\begin{array}{lrcl} \forall t \in \mathcal{T}_0: & X_t\beta + \gamma Y_{t-1} + U_t \leq & -C & \\ \forall t \in \mathcal{T}_1: & & -C & \leq X_t\beta + \gamma Y_{t-1} + U_t, \end{array}$$

the projection can be expressed as

$$\mathcal{R}_{YXU}(\theta) = \left\{(y, x, u) : \exists y_0 \in \mathcal{Y}_0 \text{ s.t. } \max_{t \in \mathcal{T}_0}\{x_t\beta + \gamma y_{t-1} + u_t\} \leq \min_{t \in \mathcal{T}_1}\{x_t\beta + \gamma y_{t-1} + u_t\}\right\}.$$

The U-level set of this projection for any (y, x) is

$$\mathcal{U}(y, x; \theta) = \left\{u : \exists y_0 \in \mathcal{Y}_0 \text{ s.t. } \max_{t \in \mathcal{T}_0}\{x_t\beta + \gamma y_{t-1} + u_t\} \leq \min_{t \in \mathcal{T}_1}\{x_t\beta + \gamma y_{t-1} + u_t\}\right\}. \tag{27}$$

In static models with $\gamma = 0$ there is the further simplification

$$\mathcal{U}(y, x; \theta) = \left\{u : \max_{t \in \mathcal{T}_0}\{x_t\beta + u_t\} \leq \min_{t \in \mathcal{T}_1}\{x_t\beta + u_t\}\right\}.$$

If observations in any periods t are missing for a class of units, for example if there is an unbalanced panel, then such t are in neither $\mathcal{T}_0$ nor $\mathcal{T}_1$ and the set $\mathcal{U}(y, x; \theta)$ leaves the value of U_t unrestricted. The identification analysis here still applies, and will result in inequalities that reflect the lack of restrictions on U_t in such periods. In a dynamic model with the value of some Y_t not observed that value becomes an additional unobserved unit-specific variable in the determination of Y_{t+1} removed *via* projection along with other such variables. Unbalanced panels can be handled in the same manner.

To illustrate identification analysis in dynamic binary response panel models consider the following example.

Example 1: Two and three period binary response.

When $T = 2$, if $Y_0 = y_0$ is observed then $\mathcal{Y}_0 = \{y_0\}$ and (27) simplifies as follows.

$$\begin{aligned}
\mathcal{U}((y_0, 0, 0), x; \theta) &= \mathcal{R}_U, \\
\mathcal{U}((y_0, 0, 1), x; \theta) &= \left\{u : u_{21}^{\Delta} \geq -x_{21}^{\Delta}\beta + \gamma y_0\right\}, \\
\mathcal{U}((y_0, 1, 0), x; \theta) &= \left\{u : u_{21}^{\Delta} \leq -x_{21}^{\Delta}\beta + \gamma(y_0 - 1)\right\}, \\
\mathcal{U}((y_0, 1, 1), x; \theta) &= \mathcal{R}_U.
\end{aligned}$$

Identification regions for θ using the inequalities arising when $Y = (0, 1)$ and when $Y = (1, 0)$ are provided by Aristodemou (2021) for models in which Y_0 is observed and either one of $U \perp\!\!\!\perp X \mid Y_0$ or $U \perp\!\!\!\perp X$ hold. Khan, Ponomareva, and Tamer (2023) provide sharp identification regions for θ in dynamic binary response models for arbitrary finite T under a conditional stationarity restriction, with Y_0 observed.

By contrast, taking the projection approach, it is not necessary to have Y_0 observed. When Y_0 is not observed the set $\mathcal{U}(y, x, \theta)$ is simply the union of the sets $\mathcal{U}(y, x; \theta)$ obtained on setting $y_0 = 0$ and then $y_0 = 1$. The sets corresponding to $Y \in \{(0, 0), (1, 1)\}$ are unchanged; the others are as follows.

$$\mathcal{U}((0, 1), x, \theta) = \left\{u : u_{21}^{\Delta} \geq -x_{21}^{\Delta}\beta - \gamma^{-}\right\}, \quad \mathcal{U}((1, 0), x, \theta) = \left\{u : u_{21}^{\Delta} \leq -x_{21}^{\Delta}\beta + \gamma^{-}\right\},$$

where $\gamma^{-} \equiv -\min\{\gamma, 0\}$. Table 1 shows the sets $\mathcal{U}(y, x; \theta)$ for the case in which $T = 3$ and Y_0 is observable.[24] Table 2 shows the sets when $T = 3$ and Y_0 is not observed.

For the case in which $T = 3$ the sets $\mathcal{U}(y, x; \theta)$ can be expressed involving just u_{31}^{Δ} and u_{32}^{Δ}, since $u_{21}^{\Delta} = u_{31}^{\Delta} - u_{32}^{\Delta}$, enabling visualization on the space of $(U_{31}^{\Delta}, U_{32}^{\Delta})$. The six nontrivial sets $\mathcal{U}(y, x; \theta)$ for each $y \in \mathcal{R}_Y$ are depicted in Figure 2 for the case in which $\theta = (\gamma, \beta)$ with $\gamma = 1$ and $X = x$ such that $-x_{32}^{\Delta}\beta = -2$, and $-x_{31}^{\Delta}\beta = 2$. $\square$

4.2 Ordered response models

Consider a dynamic ordered response panel model with

$$Y_t = \sum_{j=1}^{J} j \times 1\left[\alpha_j \leq Y_t^* < \alpha_{j+1}\right], \quad Y_t^* = X_t\beta + L_t\gamma + C + U_t, \quad t \in [T], \tag{28}$$

where $\alpha_{J+1} = -\alpha_0 = \infty$, $\theta = (\alpha_1, ..., \alpha_J, \beta, \gamma)$, and U is continuously distributed with full support on $\mathbb{R}^T$ conditional on X.[25] In the static case $\gamma = 0$.

[24] The inequalities that appear here and in similar models involving threshold crossing conditions and linear indexes are routine to derive using Fourier-Motzkin elimination.

[25] Thus as in Section 4.1 there is no loss in using weak inequalities in expressions for $\mathcal{U}(y, x; \theta)$.

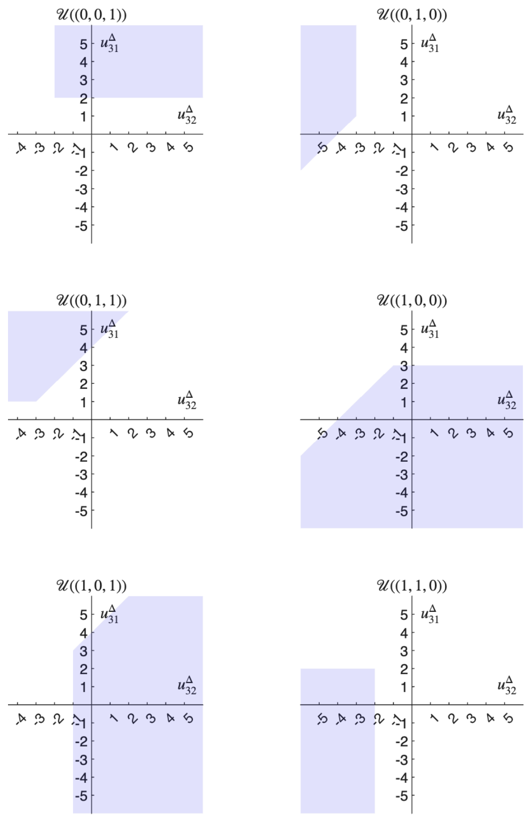


Figure 2: The support of $\mathcal{U}(Y, X; \theta)$ in a dynamic binary response model with $X = x$ and θ such that $-x_{32}^{\Delta}\beta = -2$, $-x_{31}^{\Delta}\beta = 2$, $\gamma = 1$ and unobservable initial condition, projected onto the space of $\left(U_{31}^{\Delta}, U_{32}^{\Delta}\right)$. Additionally, the sets $\mathcal{U}\left((0,0,0), x; \theta\right)$ and $\mathcal{U}\left((1,1,1), x; \theta\right)$ comprise the entire space and are therefore not illustrated.

Table 1: U-level sets in a dynamic binary response model with $T = 3$ and Y_0 observed.

y	$\mathcal{U}(y, x; \theta)$
$(0,0,0)$	$\mathcal{R}_U$
$(0,0,1)$	$\left\{u : u_{31}^{\Delta} \geq -x_{31}^{\Delta}\beta + \gamma y_0 \ \wedge \ u_{32}^{\Delta} \geq -x_{32}^{\Delta}\beta\right\}$
$(0,1,0)$	$\left\{u : u_{21}^{\Delta} \geq -x_{21}^{\Delta}\beta + \gamma y_0 \ \wedge \ u_{32}^{\Delta} \leq -x_{32}^{\Delta}\beta - \gamma\right\}$
$(0,1,1)$	$\left\{u : u_{21}^{\Delta} \geq -x_{21}^{\Delta}\beta + \gamma y_0 \ \wedge \ u_{31}^{\Delta} \geq -x_{31}^{\Delta}\beta + \gamma(y_0 - 1)\right\}$
$(1,0,0)$	$\left\{u : u_{21}^{\Delta} \leq -x_{21}^{\Delta}\beta + \gamma(y_0 - 1) \ \wedge \ u_{31}^{\Delta} \leq -x_{31}^{\Delta}\beta + \gamma y_0\right\}$
$(1,0,1)$	$\left\{u : u_{21}^{\Delta} \leq -x_{21}^{\Delta}\beta + \gamma(y_0 - 1) \ \wedge \ u_{32}^{\Delta} \geq -x_{32}^{\Delta}\beta + \gamma\right\}$
$(1,1,0)$	$\left\{u : u_{31}^{\Delta} \leq -x_{31}^{\Delta}\beta + \gamma(y_0 - 1) \ \wedge \ u_{32}^{\Delta} \leq -x_{32}^{\Delta}\beta\right\}$
$(1,1,1)$	$\mathcal{R}_U$

Table 2: U-level sets in a dynamic binary response model with $T = 3$ and Y_0 not observed.

y	$\mathcal{U}(y, x; \theta)$
$(0, 0, 0)$	$\mathcal{R}_U$
$(0, 0, 1)$	$\left\{u : u_{31}^\Delta \geq -x_{31}^\Delta\beta - \gamma^- \wedge u_{32}^\Delta \geq -x_{32}^\Delta\beta\right\}$
$(0, 1, 0)$	$\left\{u : u_{21}^\Delta \geq -x_{21}^\Delta\beta - \gamma^- \wedge u_{32}^\Delta \leq -x_{32}^\Delta\beta - \gamma\right\}$
$(0, 1, 1)$	$\left\{u : u_{21}^\Delta \geq -x_{21}^\Delta\beta - \gamma^- \wedge u_{31}^\Delta \geq -x_{31}^\Delta\beta - \gamma^+\right\}$
$(1, 0, 0)$	$\left\{u : u_{21}^\Delta \leq -x_{21}^\Delta\beta + \gamma^- \wedge u_{31}^\Delta \leq -x_{31}^\Delta\beta + \gamma^+\right\}$
$(1, 0, 1)$	$\left\{u : u_{21}^\Delta \leq -x_{21}^\Delta\beta + \gamma^- \wedge u_{32}^\Delta \geq -x_{32}^\Delta\beta + \gamma\right\}$
$(1, 1, 0)$	$\left\{u : u_{31}^\Delta \leq -x_{31}^\Delta\beta + \gamma^- \wedge u_{32}^\Delta \leq -x_{32}^\Delta\beta\right\}$
$(1, 1, 1)$	$\mathcal{R}_U$

The outcome Y_t is ordered categorical, taking the value $j \in \{0, ..., J\}$ if $Y_t^* \in [\alpha_j, \alpha_{j+1})$, where Y_t^* is a latent index. The normalization $\alpha_1 = 0$ is imposed since one of the α_j parameters can be absorbed by unit-specific variable C. The variable L_t here denotes functions of lagged outcomes in a model in which t indexes time. For example there could be $L_t = \iota_{t-1}$ where $\iota_{t-1} \equiv (1[Y_{t-1} = 1], ..., 1[Y_{t-1} = J])$ in a model with one-period lagged outcome dependence.[26] In this example $\gamma = (\gamma_1, ..., \gamma_J)'$ so $L_t\gamma = \gamma_j$ if and only if $Y_{t-1} = j$.

Now expressions for the sets of values of U that can occur given values of observed variables are derived. These are the U-level sets of the projection $\mathcal{R}_{YXU}(\theta)$ for the ordered response structural function (28).

To deal with cases in which the lag variable is not observed, define $\mathcal{L}(y, x; \theta)$ as the set of possible values of the lag variables $(L_1, ..., L_T)$ given the observability of lagged outcomes. In this exposition only one period lags are considered.

If $L_t = \iota_{t-1}$, and the initial value Y_0 is not observed, then $\mathcal{L}(y, x; \theta)$ is the set of $\iota_0, ..., \iota_{T-1}$ produced by observed $y_1, ..., y_{t-1}$ with any value of ι_0. If instead lag dependence manifests through unobservable realizations y^* as studied below, then $\mathcal{L}(y, x; \theta)$ will restrict each L_t to the interval implied by observed realizations y.

To obtain the U-level set $\mathcal{U}(y, x; \theta)$ note that there is for all s and t

$$\begin{aligned} \alpha_{y_t} - x_t\beta - l_t\gamma - u_t &\leq C \leq \alpha_{y_t+1} - x_t\beta - l_t\gamma - u_t \\ -\alpha_{y_s+1} + x_s\beta + l_s\gamma + u_s &\leq -C \leq -\alpha_{y_s} + x_s\beta + l_s\gamma + u_s \end{aligned}$$

[26] It is straightforward to accommodate multiple lags, for example two lags with $L_t = (\iota_{t-1}, \iota_{t-2})$. The indicator for one value of Y_{t-1}, here $1\{Y_{t-1} = 0\}$, is omitted by normalization as in Honoré, Muris, and Weidner (2025).

and upon adding

$$\alpha_{y_t} - \alpha_{y_s+1} - x_{ts}^{\Delta}\beta - l_{ts}^{\Delta}\gamma - u_{ts}^{\Delta} \leq 0 \leq \alpha_{y_t+1} - \alpha_{y_s} - x_{ts}^{\Delta}\beta - l_{ts}^{\Delta}\gamma - u_{ts}^{\Delta}$$

which leads to

$$\begin{aligned}\mathcal{U}(y,x;\theta) = \{u : \exists l \in \mathcal{L}\,(y,x;\theta) \text{ s.t. } \forall s < t \in [T] \\ \alpha_{y_t} - \alpha_{y_s+1} - x_{ts}^{\Delta}\beta - l_{ts}^{\Delta}\gamma \leq u_{ts}^{\Delta} \leq \alpha_{y_t+1} - \alpha_{y_s} - x_{ts}^{\Delta}\beta - l_{ts}^{\Delta}\gamma\}.\end{aligned}$$

In a static model with $\gamma = 0$ there is no lag dependence and the simplification

$$\mathcal{U}(y,x;\theta) = \{u : \forall s < t \in [T] \quad \alpha_{y_t} - \alpha_{y_s+1} - x_{ts}^{\Delta}\beta \leq u_{ts}^{\Delta} \leq \alpha_{y_t+1} - \alpha_{y_s} - x_{ts}^{\Delta}\beta\}.$$

In contrast to other approaches to dynamic ordered response panel models, unobservable initial conditions can be accommodated using the projection approach. Moreover, period t lags can include functions of both observable and latent variables, such as lagged values of the unobserved index Y^*. This is important because in many applications the ordered outcome Y_t may depend not just on the value of Y_{t-1} but on the location of Y_{t-1}^* relative to the thresholds α_j. For example, if Y_t is a categorical measure of health status, the effect of current health on future health may be different for two individuals in "good" health, one of whom is close to the boundary for the "fair" health category, and the other close to the boundary for the "excellent" health category. Or, in application to letter grades obtained in a sequence of courses, dynamic impacts may be more effectively measured by a student's numerical score rather than, say, whether they achieved an "A-" or "B+".

Lagged Outcome Dependence

With one period lagged outcome dependence $\mathcal{L}(y,x;\theta) = \mathcal{L}_1(y,x;\theta) \times \cdots \times \mathcal{L}_T(y,x;\theta)$ where $\mathcal{L}_t(y,x;\theta) = \{\iota_{t-1}\}$ for all $t > 1$, and $\mathcal{L}_1(y,x;\theta)$ is the set of standard basis vectors in $\mathbb{R}^J$.

To illustrate consider such a model with two periods, three categories, $T = J = 2$, and Y_0 not observed. When $Y = (0,0)$ or $Y = (2,2)$ any value of U is possible, as there is always C small enough or large enough, respectively, to produce either outcome. Table 3 shows the other U-level sets in this model.

Y	$\mathcal{U}(y,x;\theta)$
$(0,1)$	$\{u : u_{21}^{\Delta} \geq -x_{21}^{\Delta}\beta + \min\{0,\gamma_1,\gamma_2\}\}$
$(0,2)$	$\{u : u_{21}^{\Delta} \geq \alpha_2 - x_{21}^{\Delta}\beta + \min\{0,\gamma_1,\gamma_2\}\}$
$(1,0)$	$\{u : u_{21}^{\Delta} \leq -x_{21}^{\Delta}\beta - \gamma_1 + \max\{0,\gamma_1,\gamma_2\}\}$
$(1,1)$	$\{u : \exists y_0 \text{ s.t.} -\alpha_2 - x_{21}^{\Delta}\beta \leq u_{21}^{\Delta} - \gamma_{y_0} + \gamma_1 \leq \alpha_2 - x_{21}^{\Delta}\beta\}$
$(1,2)$	$\{u : u_{21}^{\Delta} \geq -x_{21}^{\Delta}\beta - \gamma_1 + \min\{0,\gamma_1,\gamma_2\}\}$
$(2,0)$	$\{u : u_{21}^{\Delta} \leq -\alpha_2 - x_{21}^{\Delta}\beta - \gamma_2 + \max\{0,\gamma_1,\gamma_2\}\}$
$(2,1)$	$\{u : u_{21}^{\Delta} \leq -x_{21}^{\Delta}\beta - \gamma_2 + \max\{0,\gamma_1,\gamma_2\}\}$

Table 3: U-level sets for the dynamic ordered response panel model with lagged outcome dependence, $J = T = 2$, and Y_0 not observed.

Lagged Latent Dependence

Now consider the case in which the period t outcome depends on the lagged latent index Y_{t-1}^* so $L_t = Y_{t-1}^*$. Repeated substitution for values of Y_s^*, $s < t$, in the equation for Y_t^* in (28) gives

$$Y_t^* = \gamma^t Y_0^* + \sum_{s=1}^{t} \gamma^{t-s}\big(X_s\beta + C + U_s\big).$$

So the set of values of U that deliver $Y = y$ when $X = x$ is

$$\mathcal{U}(y,x;\theta) = \left\{u : \exists (y_0^*, c) \text{ s.t. } \forall t \in [T],\ \alpha_{y_t} \leq \gamma^t y_0^* + \sum_{s=1}^{t} \gamma^{t-s}\big(x_s\beta + c + u_s\big) \leq \alpha_{y_t+1}\right\}.$$

The set $\mathcal{U}(y,x;\theta)$ differs from the case with lagged outcome dependence. The inequalities defining $\mathcal{U}(y,x;\theta)$ are linear in c and y_0^* so Fourier-Motzkin elimination can be used to remove these variables from the inequalities that define $\mathcal{U}(y,x;\theta)$.

4.3 Identified sets

With the U-level sets defined as in discrete outcome models such as those presented in Sections 4.1 and 4.2, moment inequality characterizations of identified sets for common parameters can be obtained using Artstein's inequality.[27] The inequality provides the following corollary to Proposition 1.

Corollary 1 *Suppose that Restrictions PM and RCS hold. Then the identified set*

[27] Artstein's inequality is established in Artstein (1983), see also Molchanov (2017) pages 83–84 and Molchanov and Molinari (2018) Section 2.2.

for $\left(f, G_{U|X}(\cdot|\cdot)\right)$ *comprises those pairs* $\left(f, G_{U|X}(\cdot|\cdot)\right) \in \mathcal{M}$ *such that*

$$\mathbb{P}\left[\mathcal{U}(Y,X;f) \subseteq \mathcal{S}|X=x\right] \leq G_{U|X}\left(\mathcal{S}|x\right), \text{ a.e. } x \in \mathcal{R}_X. \tag{29}$$

for all closed $\mathcal{S} \subseteq \mathcal{R}_U$. *The identified set for* f *is the set of* f *such that (29) holds for* f *and some* $G_{U|X}(\cdot|\cdot) \in \mathsf{G}_{U|X}$ *with* $\left(f, G_{U|X}(\cdot|\cdot)\right) \in \mathcal{M}$.

It will now be demonstrated how this corollary can be specialized to produce moment inequality characterizations of identified sets for common parameters. Prior applications of this inequality in the partial identification literature include Beresteanu, Molchanov, and Molinari (2012) and Chesher and Rosen (2017), see Molinari (2020) for further references. The novelty here is not in the use of Artstein's inequality for identification analysis, but rather its application to level sets of the projection $\mathcal{R}_{YXU}$ obtained by removal of incidental parameters, and under distributional restrictions commonly found in panel models having no counterpart in cross section models.

The characterization of the identified set provided by Corollary 1 using Artstein's inequality comprises for each $x \in \mathcal{R}_X$ as many inequalities as the number of closed sets in $\mathcal{R}_U$. Previous papers such as Galichon and Henry (2011), Chesher and Rosen (2017), and Luo, Ponomarev, and Wang (2026) have characterized *core determining collections* that comprise a smaller collection of sets $\mathcal{S}$ such that if (29) holds for all $\mathcal{S}$ in the collection, then it holds for all closed sets. To the best of our knowledge these results have not been previously employed in panel models but they are applicable here to simplify characterizations of identified sets delivered by Artstein's inequality.

With the characterizations of the U-level sets of Sections 4.1 and 4.2 the inequality (29) delivers observable implications for the common parameters. This is now illustrated in the context of Example 1 in Section 4.1. The same steps can be taken to characterize identified sets for the ordered response panel models of Section 4.2.

Example 1, continued: Consider the dynamic binary panel model with unobservable initial condition and $T = 3$. The U-level sets for a particular x and θ are illustrated in Figure 2. We can see immediately from the figure that with, for example, $\mathcal{S} = \left\{u : u_{31}^{\Delta} \geq u_{32}^{\Delta}\right\}$ the inequality (29) becomes $\mathbb{P}\left[Y \in \{(0,1,0),(0,1,1)\}|X=x\right] \leq G_{U|X}\left(\mathcal{S}|x\right)$. This set is however not amongst the minimal core determining collection.

From Theorem 1 of Chesher and Rosen (2017) it follows that only sets $\mathcal{S}$ that comprise unions of sets on the support of $\mathcal{U}(Y,X;f)$ need consideration. Theorem 3 of that paper establishes that among this collection, one need not consider those sets $\mathcal{S}$ that can be partitioned into two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that either $\mathcal{U}(Y,X;f) \subseteq \mathcal{S}_1$ or $\mathcal{U}(Y,X;f) \subseteq \mathcal{S}_2$, but not both simultaneously. Such a set $\mathcal{S}$ is not *self-connected* in the terminology of Luo, Ponomarev, and Wang (2026).[28] The minimal core determining collection of sets for the value of x and θ that produce the U-level sets of Figure 2 yields 32 moment inequalities of the form (29). $\square$

If U and X are stochastically independent there is the following simplification of Artstein's inequality

$$\sup_{x\in\mathcal{R}_X} \mathbb{P}\left[\mathcal{U}(Y,X;f)\subseteq\mathcal{S}|X=x\right]\leq G_U\left(\mathcal{S}\right). \tag{30}$$

This applies with both parametric and nonparametric restrictions on the class of functions f and distributions $G_U(\cdot|\cdot)$ admitted by the model. For example, if the utility function and distribution of unobservable heterogeneity are parametrically specified up to $\theta\in\Theta\subseteq\mathbb{R}^d$ in (30), f may be replaced by θ and $G_U(\mathcal{S})$ by $G_U(\mathcal{S};\theta)$.

Even when using only core determining collections of sets, the number of inequalities can be large. Nonetheless, approaches for asymptotic inference with infinitely many conditional moment inequalities can be used, see for instance Section 2.2 of Chernozhukov, Chetverikov, and Kato (2019) and Example 2 of Andrews and Shi (2017).

The next section shows how identification analysis can proceed under nonparametric specifications of the distribution of within-unit-varying heterogeneity.

4.4 Nonparametric distributional specifications

An advantage of the projection approach developed in this paper is that it enables identification analysis when there are neither parametric nor stationarity restrictions on the distribution of within-unit-varying heterogeneity. In Section 3 it was shown how this can be achieved using moment restrictions. Here are two alternative approaches more suited to models of discrete outcomes. Other such restrictions are possible.

Consider models such as the discrete outcome models considered in this section in which U level sets are determined entirely by restrictions on differences u^{Δ}_{st} for a collection of values of s and t.

First consider *quantile independence restrictions.* For chosen values of s and t,

[28]That paper also shows that it is generally possible to achieve further refinement, establishing that among the class of sets comprising unions of sets on the support of $\mathcal{U}(Y,X;f)$, those that are both self-connected and complement-connected comprise a minimal core-determining collection. However, in the panel models studied here in which $\mathcal{U}(Y,X;f)=\mathcal{R}_U$ for some $y\in\mathcal{R}_Y$, the requirement that sets be complement-connected provides no reduction in the core-determining collection.

first define an ascending sequence of quantile probabilities

$$p_{st} \equiv (p_{st}^0, p_{st}^1, \ldots, p_{st}^K, p_{st}^{K+1})$$

which are specified values in $[0,1]$ with $p_{st}^0 \equiv 0$ and $p_{st}^{K+1} \equiv 1$. Then define additional parameters, namely the unknown elements of

$$\lambda_{st} \equiv \{\lambda_{st}^0, \lambda_{st}^1, \ldots, \lambda_{st}^K, \lambda_{st}^{K+1}\}$$

with $\lambda_{st}^0 \equiv -\infty$, $\lambda_{st}^{K+1} \equiv +\infty$. These new parameters are values of the quantiles of the marginal distributions of the U_{st}^Δ at the chosen quantile probabilities, e.g. $(0.25, 0.5, 0.75)$, and there are the restrictions[29]

$$\forall x \in \mathcal{R}_X \quad \mathbb{P}[U_{st}^\Delta \leq \lambda_{st}^k | X = x] \equiv p_{st}^k \quad \text{free of } x.$$

Moment inequalities characterizing the identified set of values of the parameters θ and the quantile values in λ_{st} are

$$\forall k \in \{1, \ldots, K+1\} \quad \sup_{x \in \mathcal{R}_X} \mathbb{P}[\mathcal{U}(Y, X; \theta) \subseteq \{u : \lambda_{st}^{k-1} \leq u_{st}^\Delta \leq \lambda_{st}^k\} | X = x] \leq p_{st}^k - p_{st}^{k-1}$$

$$\forall k \in \{1, \ldots, K\} \quad \sup_{x \in \mathcal{R}_X} \mathbb{P}[\mathcal{U}(Y, X; \theta) \subseteq \{u : u_{st}^\Delta \leq \lambda_{st}^k\} | X = x] \leq p_{st}^k$$

for all pairs (s, t) for which the quantile independence restrictions are maintained. Identified sets for θ are obtained as those values of θ for which there exist values of λ_{st} such that all such inequalities are satisfied. Chesher, Kim, and Rosen (2023) gives details and has an example of this approach in action in a different, non-panel, context.[30]

Finally consider *pairwise conditional exchangeability restrictions* requiring that, for some chosen s and t, U_s and U_t are exchangeable conditional on X. Under this restriction the median of U_{st}^Δ conditional on $X = x$ is zero for all x. The inequalities above with $K = 1$, $\lambda_{st}^1 = 0$ and $p_{st}^1 = 0.5$ deliver bounds on θ absent parametric restrictions on the distribution of U.[31]

Example 1, continued: Consider again the dynamic binary outcome model as in

[29] It is easy to impose restrictions of symmetry and unimodality if that were desired.

[30] Chesher, Kim, and Rosen (2023) studies an IV Tobit model in which explanatory variables may be endogenous. Proposition 4 in Section 4.2 deals with quantile independence restrictions and is easily extended to the models considered in this paper.

[31] Under the conditional exchangeability restriction the probability density function of U_{st}^Δ is symmetric around zero. This may deliver additional bounds in some cases.

Table 4: Expressions $w_j(x, \beta, \gamma)$ in (31) required to be no greater than 1/2 under a restriction that each U_s and U_t are exchangeable conditional on X with $T = 3$.

j	$\mathcal{S}$	$w_j(x, \beta, \gamma)$
1	$\{u : u^\Delta_{31} \leq 0\}$	$p_x(1,1,0) \times 1[-x^\Delta_{31}\beta + \gamma^- \leq 0] + p_x(1,0,0) \times 1[-x^\Delta_{31}\beta + \gamma^+ \leq 0]$
2	$\{u : u^\Delta_{31} \geq 0\}$	$p_x(0,0,1) \times 1[-x^\Delta_{31}\beta - \gamma^- \geq 0] + p_x(0,1,1) \times 1[-x^\Delta_{31}\beta - \gamma^+ \geq 0]$
3	$\{u : u^\Delta_{32} \leq 0\}$	$p_x(0,1,0) \times 1[-x^\Delta_{32}\beta - \gamma \leq 0] + p_x(1,1,0) \times 1[-x^\Delta_{32}\beta \leq 0]$
4	$\{u : u^\Delta_{32} \geq 0\}$	$p_x(0,0,1) \times 1[-x^\Delta_{32}\beta \geq 0] + p_x(1,0,1) \times 1[-x^\Delta_{32}\beta + \gamma \geq 0]$
5	$\{u : u^\Delta_{21} \leq 0\}$	$(p_x(1,0,0) + p_x(1,0,1)) \times 1[-x^\Delta_{21}\beta + \gamma^- \leq 0]$
6	$\{u : u^\Delta_{21} \geq 0\}$	$(p_x(0,1,0) + p_x(0,1,1)) \times 1[-x^\Delta_{21}\beta - \gamma^- \geq 0]$

(26) with unobserved initial value Y_0 and $T = 3$ with U-level sets shown in Table 2.

The zero median independence restriction implied by pairwise conditional exchangeability of all elements of U delivers the identified set of values of (β, γ) as those satisfying

$$\bigwedge_{j=1}^{6} \left(\sup_{x \in \mathcal{R}_X} w_j(x, \beta, \gamma) \leq \frac{1}{2} \right) \tag{31}$$

where the expressions $w_j(x, \beta, \gamma)$ are shown in Table 4. In this table, $p_x(y_1, y_2, y_3) \equiv \mathbb{P}[Y = (y_1, y_2, y_3)|X = x]$ and the column headed "$\mathcal{S}$" shows the set $\mathcal{S}$ in the zero median independence restriction

$$\forall x \in \mathcal{R}_X, \quad \mathbb{P}[U \in \mathcal{S}|X = x] = 1/2$$

that delivers each row of the table. □

4.5 Related Literature on Discrete Outcome Panel Models

Analysis of binary response panel models has a long history going back to Rasch (1960, 1961), Andersen (1970), and see also Chamberlain (2010), in which a static model is studied with the elements of U restricted to be i.i.d. logistic, independent of (C, X). With these distributional restrictions β is point-identified under a rank condition and consistently estimated by a conditional maximum likelihood estimator that conditions on $\sum_{t=1}^{T} Y_t$.[32]

Extensions of panel logit models to dynamic models have been considered. Honoré and Kyriazidou (2000) provides results for identification and estimation of a dynamic binary panel model maintaining mutual independence of all elements of U and in-

[32] A precise statement of the rank condition is provided as Assumption 2 in Davezies, D'Haultfœuille, and Laage (2024).

dependence of U and (C, X), and in most cases restricting the elements of U to be logistically distributed. Kitazawa (2022), Dano (2023), and Honoré and Weidner (2025) provide moment equations in dynamic panel logit models, which can be used to study identification and estimation of common parameters. Dobronyi, Gu, Kim, and Russell (2025) analyze the full likelihood from the dynamic panel logit model and make a connection to the truncated moment problem to obtain all of the model's observable implications. That paper and Davezies, D'Haultfœuille, and Laage (2024) also provide characterizations of certain average and marginal effects.

An alternative to these logit specifications in the binary outcome panel model is a conditional *stationarity* restriction introduced in Manski (1987), requiring that conditional on (X, C) the variables $U_1, ..., U_T$ all have the same marginal distribution. This is implied by the panel logit distributional restriction, but is weaker. It does not require independence of U and (X, C), and it can allow for correlation in the components of U. Nonetheless it does restrict the joint distribution of U and C.[33]

Conditional stationarity restrictions have been used in several papers. Abrevaya (2000) studies a class of generalized regression models that nests binary response and censored outcome models under conditional stationarity and stronger restrictions. Chernozhukov, Fernandez-Val, Hahn, and Newey (2013) characterizes bounds for average and quantile effects in several nonseparable panel models, including binary response models. Khan, Ponomareva, and Tamer (2023) provides set identification results for common parameters in semiparametric dynamic binary response panel models. Conditional stationarity restrictions have also been used in multinomial response panel models, for example in Shi, Shum, and Song (2018), Khan, Ouyang, and Tamer (2021), Pakes and Porter (2024), Pakes, Porter, Shepard, and Calder-Wang (2025), Gao and Wang (2026), Gao and Li (2026), and Mbakop (2023).

Aristodemou (2021) is the one paper of which we are aware that studies the binary response specification (26) without restricting the covariation of C with either X or U. In that paper U and X are restricted to be independently distributed, in some cases conditional on an initial condition, but, importantly, not conditional on C. That paper provides bounds on parameters in the models studied but does not claim sharpness. The projection approach delivers characterizations of sharp identified sets and applies more broadly, for example allowing arbitrary T, unobserved initial conditions, and alternative restrictions on the joint distribution of U and X.

The literature on fixed effects models of dynamic ordered response panels is recent

[33]As pointed out in Chernozhukov, Fernandez-Val, Hahn, and Newey (2013) the stationarity restriction $U_t|C, X \stackrel{d}{=} U_1|C, X$ for all t is equivalent to $(U_t, C)|X \stackrel{d}{=} (U_1, C)|X$ for all t.

and not extensive. Honoré, Muris, and Weidner (2025) employ functional differencing to provide moment conditions that can be used as a basis for estimation and inference in models in which the elements of U are i.i.d. logistically distributed and independent of X and C. In a model with these distributional restrictions on U but an alternative lag dependence specification $L_t = 1[Y_{t-1} \geq k]$ for specified k, Muris, Raposo, and Vandoros (2025) develop a conditional maximum likelihood estimator building on insights from Honoré and Kyriazidou (2000). References to the broader literature on ordered response panel models, including random effects approaches and static models, can be found in these papers. An unpublished chapter of Aristodemou (2016) provides bounds on common parameters in some fixed effects ordered response panel models when $T = 2$.[34]

5 Numerical Illustrations of Identified Sets

Illustrations of identified sets for (β, γ) are presented for the CES model of Section 3.1. Both weak and strict exogeneity restrictions, MW and MS, are considered.

As in Section 3.1, the model features firm-specific unobservables C and S in the CES production function. For the sake of illustration, a data generation process is considered in which $T = 3$, and, although it is unknown to the econometrician, $C = S = 0$ for all firms, so output follows a Cobb-Douglas specification

$$Y_t = \beta_0 \left(\gamma_0 \log(X_{t1}) + (1 - \gamma_0) \log(X_{t2})\right) + U_t, \qquad t \in \{1, 2, 3\}.$$

This is chosen for simplicity, but calculations are easily done for more complex cases.

Let X denote the 3x2 matrix with elements X_{tj}. To determine the support of X, values of its elements X_{t1} and X_{t2} were drawn i.i.d. with $\log X_{tj} \sim N(0, 1/4)$ to produce 50 support points, each of which was given equal probability.

[34] The models studied in Chapter 6 of Aristodemou (2016), like those studied in this paper, impose no restrictions on the joint distribution of C and U, although the analysis requires an observed initial condition, independence restrictions conditional on the initial condition, and does not allow dependence on lagged latent variables Y^* as is allowed here. Moreover, non-sharp outer bounds are obtained. However, in contrast to Honoré and Kyriazidou (2000) and Muris, Raposo, and Vandoros (2025), in Aristodemou (2016), as here, no logistic or other parametric distributional restriction on U is required.

The identified set is given by the inequalities (25), equivalently for each r:

$$E\left[\max_{s\in[-\infty,1]}\sum_{t=1}^{T} w_t(r,X)\left(Y_t-\beta\log g(X_{t1},X_{t2},\gamma,s)\right)\right]-E\left[\inf_{c\in\mathbb{R}}\sum_{t=1}^{T} w_t(r,X)c\right]\geq 0. \tag{32}$$

For r such that $\Pr\left[\sum_{t=1}^{T} w_t(r,X)=0\right]<1$, the last term can be made arbitrarily large and (32) will be satisfied for any such r. Therefore we have the characterization

$$\min_{r:\|r\|=1} E\left[\max_{s\in[-\infty,1]}\sum_{t=1}^{T} w_t(r,X)\left(Y_t-\beta\log g(X_{t1},X_{t2},\gamma,s)\right)\right]\geq 0, \tag{33}$$

for which we need only consider values of r such that $\sum_{t=1}^{T} w_t(r,X)=0$ almost surely. If the support of X is such that there exists no proper linear subspace of $\mathbb{R}^{2T+1}$ that contains $(1,X_1,...,X_T)$ almost surely, as is the case in this illustration, this is equivalent to imposing the restrictions

$$\sum_{t=1}^{T} r_t=0 \quad\text{and}\quad \sum_{t=1}^{T} r_{thk}=0,\ \text{for all } h,k. \tag{34}$$

This has the effect of removing c from the inequality.

On replacing Y_t by its expectation conditional on X, denoted $\mu_t(X)$, there is:

$$\min_{r\in\mathcal{R}} E\left[\max_{s\in[-\infty,1]}\sum_{t=1}^{T} w_t(r,X)\Big(\mu_t(X)-\beta\log g(X_{t1},X_{t2},\gamma,s)\Big)\right]\geq 0, \tag{35}$$

where $\mathcal{R}$ is the set of r such that $\|r\|=1$ and (34) holds.

Weak and strict exogeneity restrictions are distinguished by additionally imposing $r_{thk}=0$ for all $t<h$ under weak exogeneity, as described in Section 3.3.1.

The maximisation with respect to s is done using the modified golden section method provided by the `optimise` function of R.[35] Maximisation is done with respect to two alternative monotone transformations of $s\in(-\infty,1]$ to the unit interval:

$$\lambda_1(s)=\exp(s-1), \qquad \lambda_2(s)=\frac{1}{\pi}\arctan(-\log(1-s))+\frac{1}{2}.$$

Any internal maxima that are found are compared with the values obtained at $s=-\infty$ and $s=1$ and the largest value is chosen. The expectation in (35) is obtained as the probability-weighted sum over the support of X.

[35] R Core Team (2025): `https://www.R-project.org/`.

Under the strict exogeneity Restriction MS stated in Section 3.1, with $T = 3$ there are 21 elements in r subject to $\|r\| = 1$ and the seven restrictions (34). Under the weak exogeneity Restriction MW there are an additional six restrictions $r_{thk} = 0$ for $t < h$.[36] Even in the case with r more heavily restricted, minimization is hard to calculate precisely so an alternative calculation is done that delivers an outer region. The bounds obtained nonetheless demonstrate the informativeness of the CES model with multiple fixed effects, one entering nonlinearly.

The calculation proceeds by drawing 5000 pseudo-random standard Gaussian values of r which are subjected to the required restrictions. A value of (β, γ) is deemed out of the identified set if the inequality (35) is violated at any of the values of r considered. The same stream of 5000 pseudo-random values of r are employed as each value of (β, γ) is considered.

Figure 3 shows the results obtained under the weak (dark blue) and strict (light blue) exogeneity restrictions when $(\beta_0, \gamma_0) = (1, 0.5)$. These outer regions are quite informative even in this simple case in which $T = 3$. Having larger T or taking more than 5000 pseudo-random draws would deliver tighter bounds.

6 Discussion and concluding remarks

In the econometrics and the statistics literature the incidental parameter problem arising with short panels has mostly been subject to analysis using particular parametric specifications of distributions of outcomes or unobservables. Notable examples are Neyman and Scott (1948), Honoré and Kyriazidou (2000), Lancaster (2000), and Bonhomme (2012). A problem for practicing researchers is: which distribution to choose - economic reasoning and context usually offers little guidance, and the literature says little about the consequences of an unsuitable choice.

A notable exception is the work based on the stationarity restrictions introduced in Manski (1987). In that work no parametric restrictions are placed on probability distributions, but the approach is not universally applicable.

The situation in the year 2000 was summarized by Tony Lancaster as follows:[37] "The absence of a method guaranteed to work in a large class of econometric models means that any paper on [the incidental parameter problem] must be a catalogue of examples". Little has changed in the years that followed. The projection approach

[36] Note that imposing $r_{thk} = 0$ for $t < h$ in (35) corresponds to a less restrictive model than when $r_{thk} = 0$ for $t < h$ is not imposed.

[37] Lancaster (2000), page 404.

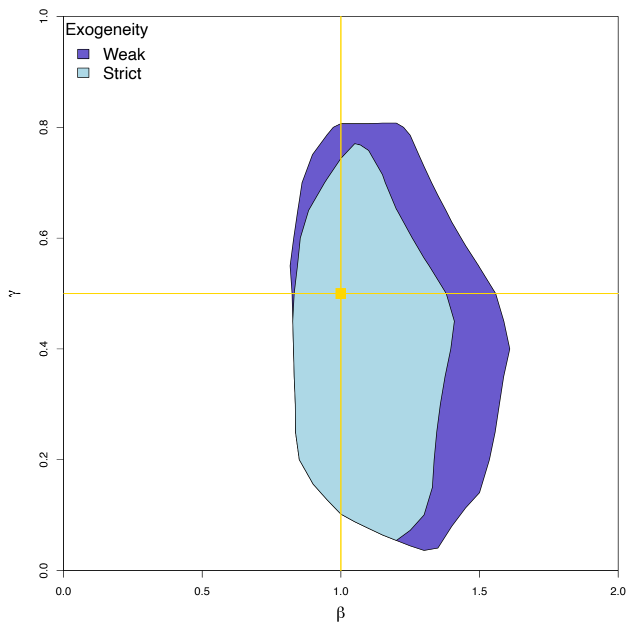


Figure 3: Bounds on (β, γ) in the CES Model for the data generation process described in Section 5 under weak and strict exogeneity restrictions. The lines and point plotted in yellow show the values of β and γ in the process generating the distribution of (Y, X) used in this illustration.

introduced in this paper fills this gap, delivering a universally applicable solution to the incidental parameter problem.

Taking this projection approach, incidental parameters in any number are projected away from the space of observed and unobserved variables. The original model specification then delivers correspondences specifying the feasible combinations of the remaining variables. Classical "fixed effects", unobserved initial conditions and missing data are examples of variables that can be treated in this way.

Projection delivers an incomplete model whose identifying power can be determined by extension of available methods, for example as developed for the analysis of Generalized Instrumental Variable models in Chesher and Rosen (2017). Estimation and inference using the resulting characterizations of identified sets is off-the-shelf.

With the incidental parameters projected away, robust econometric analysis can proceed absent restrictions on their joint probability distribution with other variables. Importantly, progress can be made using nonparametric specifications of the distribution of the unobserved variables that vary within observational units. For example, mean and conditional mean restrictions can be employed as in the CES production function example of Section 3 and quantile independence restrictions can be used as described in Section 4.

Four examples of application to econometric models have been set out in this paper. More can be found in the online working paper Chesher, Rosen and Zhang (2024), which includes applications to models admitting multiple indexes, for example, models of multiple discrete choice and simultaneous binary response.

Endogenous explanatory variables are easily accommodated following the GIV analysis of Chesher and Rosen (2017). All endogenous variables are placed in the list of outcomes, Y, and restrictions suitable for the context are imposed on the distribution of within-observation-unit-varying U and explanatory variables, X. It is straightforward to impose *weak* exogeneity restrictions, for example, in dynamic panels requiring that for all t, $(X_1, \ldots, X_t)$ and $(U_t, \ldots, U_T)$ satisfy some suitable-for-context independence restriction while allowing feedback from historic shocks and outcomes to affect the determination of future X values.

Finally, the results of this paper can be useful for conducting sensitivity analysis and specification testing. The identified sets delivered by this paper's models that place no restriction on the distribution of unobservable unit-specific effects will contain the structures identified by more restrictive models if their restrictions are satisfied by the process under study, as captured in the distribution of observable variables the process delivers. The analysis set out here can show how sensitive the findings obtained using those more restrictive models are to relaxation of their additional restrictions. It may be found that estimation employing a point-identifying model delivers a structure outside an estimator of the identified set obtained using a less restrictive model of the type studied in this paper. That will suggest the more restrictive model is misspecified. Formal development of such specification tests is a potentially fruitful topic for future research.

References

Abrevaya, Jason (2000), "Rank estimation of a generalized fixed effects regression model." *Journal of Econometrics*, 95 (1), 1–23.

Abrevaya, Jason and Chris Muris (2020), "Interval censored regression with fixed effects." *Journal of Applied Econometrics*, 35 (2), 198–216.

Andersen, Erling Bernhard (1970), "Asymptotic properties of conditional maximum-likelihood estimators." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 32 (2), 283–301.

Andrews, Donald W. K. and Xiaoxia Shi (2017), "Inference based on many conditional moment inequalities." *Journal of Econometrics*, 196 (2), 275–287.

Arellano, Manuel and Stephen Bond (1991), "Some tests of specification for panel data models: Monte carlo evidence and an application to unemployment equations." *Review of Economic Studies*, 58 (2), 277–297.

Aristodemou, Eleni (2016), "Essays on econometric models of discrete choice for vertically differentiated alternatives in oligopoly models." PhD Dissertation, University College London.

Aristodemou, Eleni (2021), "Semiparametric identification in panel data discrete response models." *Journal of Econometrics*, 220 (2), 253–271.

Artstein, Zvi (1983), "Distributions of random sets and random selections." *Israel Journal of Mathematics*, 46, 313–324.

Bai, Yuehao, Andres Santos, and Azeem M Shaikh (2022), "A two-step method for testing many moment inequalities." *Journal of Business and Economic Statistics*, 40 (3), 1070–1080.

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79 (6), 1785–1821.

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari (2012), "Partial identification using random set theory." *Journal of Econometrics*, 166 (1), 17–32.

Bonhomme, Stephane (2012), "Functional differencing." *Econometrica*, 80 (4), 1337–1385.

Bonhomme, Stephane (2025), "Back to feedback. dynamics and heterogeneity in panel data." Working paper.

Bonhomme, Stéphane, Kevin Dano, and Bryan Graham (2023), "Identification in a binary choice panel data model with a predetermined covariate." *arXiv preprint arXiv:2301.05733*.

Bonhomme, Stéphane, Kevin Dano, and Bryan Graham (2025), "Moment restrictions for nonlinear panel data models with feedback."

Canay, Ivan A, Gaston Illanes, and Amilcar Velez (2026), "A user's guide for inference in models defined by moment inequalities." *Journal of Econometrics*. Forthcoming.

Chamberlain, Gary (2010), "Binary response models for panel data: Identification and information." *Econometrica*, 78 (1), 159–168.

Chamberlain, Gary (2022), "Feedback in panel data models." *Journal of Econometrics*, 226 (1), 4–20.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2019), "Inference on causal and structural parameters using many moment inequalities." *Review of Economic Studies*, 86 (5), 1867–1900.

Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey (2013), "Average and quantile effects in nonseparable panel models." *Econometrica*, 81 (2), 535–580.

Chesher, Andrew, Dongwoo Kim, and Adam Michael Rosen (2023), "IV methods for tobit models." *Journal of Econometrics*, 235 (2), 1700–1724.

Chesher, Andrew and Adam M. Rosen (2015), "Characterizations of identified sets delivered by structural econometric models." CeMMAP working paper CWP63/15.

Chesher, Andrew and Adam M Rosen (2017), "Generalized instrumental variable models." *Econometrica*, 85 (3), 959–989.

Chesher, Andrew and Adam M Rosen (2020), "Generalized instrumental variable models, methods, and applications." In *Handbook of Econometrics*, Vol. 7, 1–110, Elsevier.

Dano, Kevin (2023), "Transition probabilities and moment restrictions in dynamic fixed effects logit models." *arXiv preprint arXiv:2303.00083*.

Davezies, Laurent, Xavier D'Haultfœuille, and Louise Laage (2024), "Identification and estimation of average marginal effects in fixed effects logit models." *arXiv preprint arXiv:2105.00879*.

Dobronyi, Christopher R., Jiaying Gu, Kyoo il Kim, and Thomas M Russell (2025), "Identification of dynamic panel logit models with fixed effects." *arXiv preprint arXiv:2104.04590*.

Galichon, Alfred and Marc Henry (2011), "Set identification in models with multiple equilibria." *Review of Economic Studies*, 78 (4), 1264–1298.

Gao, Wayne Yuan and Ming Li (2026), "Identification of semiparametric panel multinomial choice models with infinite-dimensional fixed effects." *Review of Economics and Statistics*. Forthcoming.

Gao, Wayne Yuan and Rui Wang (2026), "Identification in nonlinear dynamic panels under partial stationarity." *Journal of Econometrics*, 253. 106185.

Honoré, Bo E. (1992), "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects." *Econometrica*, 60 (3), 533–565.

Honoré, Bo E. (1993), "Orthogonality conditions for tobit models with fixed effects and lagged dependent variables." *Journal of Econometrics*, 59 (1–2), 35–61.

Honoré, Bo E and Ekaterini Kyriazidou (2000), "Panel data discrete choice models with lagged dependent variables." *Econometrica*, 68 (4), 839–874.

Honoré, Bo E, Chris Muris, and Martin Weidner (2025), "Dynamic ordered panel logit models." *Quantitative Economics*, 16 (3), 899–945.

Honoré, Bo E and Elie Tamer (2006), "Bounds on parameters in panel dynamic discrete choice models." *Econometrica*, 74 (3), 611–629.

Honoré, Bo E and Martin Weidner (2025), "Dynamic ordered panel logit models." *Review of Economic Studies*, 92 (5), 3112–3137.

Hu, Luojia (2002), "Estimation of a censored dynamic panel data model." *Econometrica*, 70 (2), 2499–2517.

Khan, S, M Ponomareva, and E Tamer (2023), "Identification of dynamic binary response models." *Journal of Econometrics*, 237 (1), 105515.

Khan, Shakeeb, Fu Ouyang, and Elie Tamer (2021), "Inference on semiparametric multinomial response models." *Quantitative Economics*, 12 (3), 743–777.

Khan, Shakeeb, Maria Ponomareva, and Elie Tamer (2016), "Identification of panel data models with endogenous censoring." *Journal of Econometrics*, 194 (1), 57–75.

Kitazawa, Yoshitsugu (2022), "Transformations and moment conditions for dynamic fixed effects logit models." *Journal of Econometrics*, 229 (2), 350–362.

Lancaster, Tony (2000), “The incidental parameter problem since 1948.” *Journal of Econometrics*, 95 (2), 391–413.

Lee, Wooyong (2026), “Identification and estimation of dynamic random coefficients models.” *Review of Economic Studies*. Forthcoming.

Li, Lixiong (2026), “Identification and counterfactual analysis in incomplete models with support and moment restrictions.” ArXiv working paper 2603.07722.

Luo, Ye, Kirill Ponomarev, and Hai Wang (2026), “Selecting inequalities for sharp identification in models with set-valued predictions.” *Review of Economic Studies*. Forthcoming.

Manski, Charles F (1987), “Semiparametric analysis of random effects linear models from binary panel data.” *Econometrica*, 55 (2), 357–362.

Manski, Charles F (1988), “Identification of binary response models.” *Journal of the American Statistical Association*, 83 (403), 729–738.

Marcoux, Mathieu, Thomas M Russell, and Yuanyuan Wan (2024), “A simple specification test for models with many conditional moment inequalities.” *Journal of Econometrics*, 242 (1).

Mbakop, Eric (2023), “Identification in some discrete choice models: A computational approach.” *arXiv preprint arXiv:2305.15691*.

Molchanov, Ilya and Francesca Molinari (2018), *Random Sets in Econometrics*, Cambridge University Press, Cambridge, U.K.

Molchanov, Ilya S. (2017), *Theory of Random Sets, 2nd Edition*, Springer Verlag, London.

Molinari, Francesca (2020), “Microeconometrics with partial identification.” In *The Handbook of Econometrics* (Steven Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa Matzkin, eds.), Vol. 7a, 355–486, Elsevier.

Muris, Chris, Pedro Raposo, and Sotiris Vandoros (2025), “A dynamic ordered logit model with fixed effects.” *Review of Economics and Statistics*, 107 (4), 1104–1114.

Neyman, Jerzy and Elizabeth L Scott (1948), “Consistent estimates based on partially consistent observations.” *Econometrica*, 16 (1), 1–32.

Pakes, Ariel and Jack Porter (2024), "Moment inequalities for multinomial choice with fixed effects." *Quantitative Economics*, 15 (1), 1–25.

Pakes, Ariel, Jack Porter, Mark Shepard, and Sophie Calder-Wang (2025), "Unobserved heterogeneity, state dependence, and health plan choices." Working paper.

R Core Team (2025), *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rasch, Georg (1960), *Studies in Mathematical Psychology: 1. Probabilistic models for some intelligence and attainment tests.*, Nielsen and Lydiche.

Rasch, Georg (1961), "On general laws and the meaning of measurement in psychology." In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 4, 321–333.

Rockafellar, Tyrrell R. (1970), *Convex Analysis*, Princeton University Press, New York.

Schennach, Susanne (2014), "Entropic latent variable integration via simulation." *Econometrica*, 82 (1), 345–385.

Shi, Xiaoxia (2025), "Inference in models defined by infinitely many moment inequalities: A survey." *prepared for 2025 Econometric Society World Congress monograph.*

Shi, Xiaoxia, Matthew Shum, and Wei Song (2018), "Estimating semi-parametric panel multinomial choice models using cyclic monotonicity." *Econometrica*, 86 (2), 737–761.

A Proofs

Proof of Proposition 1. *First to be shown is that the definitional expression of* $\mathcal{I}(\mathcal{M}, F_{YX})$ *in* (3) *is equivalent to the expression in* (8). *Let* $\left(f, G_{U|X}(\cdot|\cdot)\right)$ *be a member of the set defined in* (3). *Then there exist* $\tilde{U}$ *and* $\tilde{V}$ *and* $\tilde{Y}$ *such that for almost every* $x \in \mathcal{R}_X$*, conditional on* $X = x$*: (i)* $\tilde{U} \sim G_{U|X}(\cdot|x)$*, (ii)* $\tilde{Y} \in \mathcal{Y}(\tilde{U}, \tilde{V}, X; f)$*, and (iii)* $\tilde{Y} \sim F_{Y|X}(\cdot|x)$. *Condition (ii) implies that* $\tilde{Y} \in \mathcal{Y}(\tilde{U}, X; f)$. *This with (i) and (iii) implies that* $\left(f, G_{U|X}(\cdot|\cdot)\right)$ *is an element of the set defined in* (8).

Now for the other direction let $\left(f, G_{U|X}(\cdot|\cdot)\right)$ *be an element of the set defined in* (8). *Then there exist* $\tilde{Y}$ *and* $\tilde{U}$ *such that for almost every* $x \in \mathcal{R}_X$*, conditional on*

$X = x$: (i) $\tilde{Y} \sim F_{Y|X}(\cdot|x)$, and (ii) $\tilde{Y} \in \mathcal{Y}(\tilde{U}, X; f)$, where $\tilde{U}|X = x \sim G_{U|X}(\cdot|x)$. Condition (ii) implies that for each realization of $\tilde{U}$ there exists $v \in \mathcal{R}_V$ such that $\tilde{Y} \in \mathcal{Y}(\tilde{U}, v, X; f)$. This further implies that there exists a random variable $\tilde{V}$ that assigns mass to such values of v, i.e. $\tilde{Y} \in \mathcal{Y}(\tilde{U}, \tilde{V}, X; f)$. Thus $\left(f, G_{U|X}(\cdot|\cdot)\right)$ is in the set defined by (3)*, completing the proof that* (3) *and* (8) *are equivalent characterizations of $\mathcal{I}(\mathcal{M}, F_{YX})$. The equivalence of* (8) *and* (9) *follows directly from relation* (7) *and the definition of selectionability.* ∎

Lemma 1 *Let Restriction RCS hold. For any $\left(f, G_{U|X}(\cdot|\cdot)\right) \in \mathcal{M}$, random vector $\tilde{U}$ defined on $(\Omega, \mathsf{L}, \mathbb{P})$ with conditional distribution $G_{U|X}(\cdot|x)$ for almost every x is a measurable selection of $\mathcal{U}(Y, X; f)$ if and only if the same random vector $\tilde{U}$ is a measurable selection of $\mathsf{cl}\left(\mathcal{U}(Y, X; f)\right)$.*

Proof of Lemma 1. *It is immediate that if $\tilde{U}$ is a measurable selection of $\mathcal{U}(Y, X; f)$ then it is also a measurable selection of $\mathsf{cl}\left(\mathcal{U}(Y, X; f)\right)$, so only the reverse implication needs to be shown. Thus, suppose that $\tilde{U}$ is a measurable selection of $\mathsf{cl}\left(\mathcal{U}(Y, X; f)\right)$. Under Restriction RCS the event that $\tilde{U}$ is an element of $\mathsf{cl}\left(\mathcal{U}(Y, X; f)\right)$ but not an element of $\mathcal{U}(Y, X; f)$ occurs with zero probability. Since $\tilde{U}$ is a measurable selection of $\mathsf{cl}\left(\mathcal{U}(Y, X; f)\right)$ we have that $\tilde{U} \in \mathsf{cl}\left(\mathcal{U}(Y, X; f)\right)$ almost surely. It follows that $\tilde{U} \in \mathcal{U}(Y, X; f)$ almost surely and thus $\tilde{U}$ is a measurable selection of $\mathcal{U}(Y, X; f)$.* ∎

Proof of Corollary 1. *This follows directly from Proposition 1 and Lemma 1, together with Corollary 1 of Chesher and Rosen (2017).* ∎

Proof of Proposition 2. *First it will be shown that the identified set for f is $\mathcal{F}_I = \{f \in \mathcal{F} : \mathbf{0} \in \mathbb{E}_I\left[\mathcal{Q}(Y, X; f)\right]\}$. Under Restriction M, $\mathcal{Q}(Y, X; \theta)$ has an integrable selection and is closed, so it is an integrable random closed set, and there exists random vector Q such that $Q \in \mathcal{Q}(Y, X; f)$ almost surely, i.e. a measurable selection of $\mathcal{Q}(Y, X; f)$, with mean $\mathbf{0}_J$ if and only if $\mathbf{0}_J \in \mathbb{E}_I\left[\mathcal{Q}(Y, X; f)\right]$. By definition of $\mathcal{Q}(Y, X; f)$, the set of measurable selections of $\mathcal{Q}(Y, X; f)$ is the set of random vectors $(Z_1'\tilde{U}_1, ..., Z_T'\tilde{U}_T)$ such that $\tilde{U}$ is a measurable selection of $\mathcal{U}(Y, X; f)$. Thus there exists $Q \in \mathcal{Q}(Y, X; f)$ almost surely with $E[Q] = \mathbf{0}_J$ if and only if there exists $\tilde{U} \in \mathcal{U}(Y, X; f)$ almost surely with $E[Z_t'\tilde{U}_t] = \mathbf{0}_{J_t}$ for all t. Thus $\mathcal{F}_I$ is the identified set for f.*

The moment closure of $\mathcal{F}_I$ is $\overline{\mathcal{F}_I} = \{f \in \mathcal{F} : \mathbf{0} \in \mathbb{E}\left[\mathcal{Q}(Y, X; f)\right]\}$, where $\mathbb{E}\left[\mathcal{Q}(Y, X; f)\right] = \mathsf{cl}\left(\mathbb{E}_I\left[\mathcal{Q}(Y, X; f)\right]\right)$. Under Restriction PM, $(\Omega, \mathsf{L}, \mathbb{P})$ is nonatomic and by Theorem 2.1.26 of Molchanov (2017) we have that $\mathbb{E}\left[\mathcal{Q}(Y, X; \theta)\right]$ is convex. Because

$\mathbb{E}[\mathcal{Q}(Y,X;f)]$ is convex, $\mathbf{0}_J \in \mathbb{E}[\mathcal{Q}(Y,X;f)]$ if and only if $0 \leq h(\mathbb{E}[\mathcal{Q}(Y,X;\theta)], r)$ for all $r \in \mathbb{R}^J$. Then (23) *follows because the support function is positive homogeneous and $h(\mathbb{E}[\mathcal{Q}(Y,X;\theta)], r) = E[h(\mathcal{Q}(Y,X;\theta), r)]$ by Theorem 2.1.35 of Molchanov (2017). Finally, if $\mathbb{E}_I[\mathcal{Q}(Y,X;f)]$ is closed then it is equal to $\mathbb{E}[\mathcal{Q}(Y,X;f)]$ and consequently $\mathcal{F}_I = \overline{\mathcal{F}_I}$.* ∎

Proposition 3 *Let the restrictions of Proposition 2 hold with the CES specification in* (10)*. Suppose in addition that*

$$\forall t \in [T], \quad E\left[\sup_{s\in[-\infty,1]} \|Z_t\left(Y_t - \beta \log g(X_{t1}, X_{t2}, \gamma, s)\right)\|\right] < \infty. \tag{36}$$

Then $\mathbb{E}_I[\mathcal{Q}(Y,X;\theta)]$ is closed, and the moment closure of the identified set, Θ^*, is sharp.

Proof of Proposition 3. *By definition*

$$\mathbb{E}_I[\mathcal{Q}(Y,X;\theta)] = \left\{q \in \mathbb{R}^J : q = E[Q] \text{ for some } Q \in \mathbf{L}^1(\mathcal{Q}(Y,X;\theta))\right\}.$$

Let $A_t(\theta,s) \equiv Z_t\left(Y_t - \beta \log g(X_{t1}, X_{t2}, \gamma, s)\right)$, $A(\theta,s) \equiv (A_1(\theta,s), ..., A_T(\theta,s))$, and J-element vector $Z \equiv (Z_1, ..., Z_T)$. Then $Q \in \mathbf{L}^1(\mathcal{Q}(Y,X;\theta))$ implies that for some measurable (S,C) with support in $[-\infty,1] \times \mathbb{R}$, $Q = A(\theta,S) - CZ$. Since $E[A(\theta,S)]$ is bounded by (36), $E[Q] = E[A(\theta,S)] - E[CZ]$, and $\mathbb{E}_I[\mathcal{Q}(Y,X;\theta)] = \mathcal{E}_1 + \mathcal{E}_2$ where

$$\begin{aligned}
\mathcal{E}_1 &= \{E[A(\theta,S)] : S \in [-\infty,1] \text{ and } S \text{ measurable}\} = \mathbb{E}_I[\{A(\theta,s) : s \in [-\infty,1]\}], \\
\mathcal{E}_2 &= \{E[-CZ] : E[|C|\|Z\|] < \infty \text{ and } C \text{ measurable}\}.
\end{aligned}$$

The set $\{A(\theta,s) : s \in [-\infty,1]\}$ is an integrably bounded random compact set by (36), so $\mathcal{E}_1$ is compact by Theorem 2.1.38 of Molchanov (2017). Both $\mathcal{E}_1$ and $\mathcal{E}_2$ are convex because the underlying probability space is nonatomic by Theorem 2.1.26 of Molchanov (2017). The set $\mathcal{E}_2$ is a linear subspace of $\mathbb{R}^J$ and is therefore closed. The sum of a compact convex set and a closed convex set is closed, see e.g. Corollary 9.1.2 of Rockafellar (1970), so $\mathbb{E}_I[\mathcal{Q}(Y,X;\theta)]$ is closed and equal to $\mathbb{E}[\mathcal{Q}(Y,X;\theta)]$, completing the proof. ∎