SEMIPARAMETRIC EFFICIENT ESTIMATION IN STRUCTURAL TIME SERIES CROSS SECTION MODELS

IVANA KOMUNJER AND GIUSEPPE RAGUSA

Abstract. In this paper we study a question of semiparametric efficiency bounds for finite dimensional parameters of structural time series cross section models. When the models exhibit both temporal and spatial dependence, and heterogeneity among variables, little is known about this question. Our contribution is twofold. First, we construct a (least favorable) parametric submodel of a semiparametric model defined by a panel exogeneity condition. This condition—not seen in previous work—generalizes the notion of strict stationarity used in setups with independent and identically distributed variables. Second, we derive the asymptotic distribution of the maximum likelihood estimator obtained in the (least favorable) submodel, thus providing a tight lower bound on the semiparametric efficiency.

[Preliminary and Incomplete. Please do not circulate.]

Affiliations and contact information. Komunjer: Department of Economics, University of California at San Diego, komunjer@ucsd.edu. Ragusa: Department of Economics, University of California at Irvine, gragusa@uci.edu.

Acknowledgments. The authors wish to thank the participants in NBER/NSF 2007 Time Series conference in Iowa City, as well as in Econometrics seminars at Toulouse, UCI, UCLA, UCR, UCSD, UC Davis, U Michigan and MSU for their suggestions and comments. Special thanks to Bruce Hansen, Nour Meddahi, Peter Schmidt, Jim Stock, and Jeff Wooldridge.

1. Introduction

This paper considers structural cross section time series models that come in a form of a system of dynamic equations. The equations are allowed to be nonlinear, both in variables and in parameters. The use of nonlinearities may be necessary to represent certain features of the underlying economic model, such as: dynamic behavior (e.g. dynamic adjustments), the form of a utility or production function, the institutional or market structure.

Specifically, we focus on structural models that are finitely parameterized, and the finite dimensional structural parameter is the focal point of the estimation problem. The models are semiparametric in nature: the random variables which they involve are only known to satisfy a certain number of conditional moment restrictions; their joint distribution is, however, left unspecified. This raises a question of semiparametric efficiency bound for the structural parameter of interest.

There are several important antecedents that deal with semiparametric efficiency in a special—linear—case of our structural cross section time series model: that of dynamic panels. Works by Chamberlain (1992), Hayashi (1992), Keane and Runkle (1992), Schmidt et al. (1992), Arellano and Bover (1995), Ahn and Schmidt (1995) and Park et al. (2007), for example, address the question of semiparametric efficient estimation under under various exogeneity assumptions relating the explanatory variables and the unobserved effects.

Common to all of the above work is a strong assumption on the cross-sectional structure of the variables in the panel: while allowed to be correlated in the time direction, they remain *independent and identically distributed across individuals*. In other words, the existing results on semiparametric efficiency apply to dynamic panel models that do not allow for any heterogeneity nor cross section dependence. As pointed out by Baltagi and Pesaran (2007), various spatial or spill over effects, as well as unobserved (or unobservable) common factors, can result in unobserved errors that are heterogeneous and dependent across individuals. While well recognized by the recent literature on non-stationary panel data, the problem of heterogeneity and cross section dependence has remained unsolved by the research on semiparametric efficiency. The main contribution of our paper is to fill this gap.

Specifically, we derive the semiparametric efficiency bound for the finite dimensional structural parameter under a *panel exogeneity* condition, not yet seen in the literature. The notion of panel exogeneity extends that of strict exogeneity to cross section time series structural systems that allow for dependence both across time as well as individuals.

We now give an insight into the paper's key result by examining a well-known dynamic panel model:

$$Y_{t,i} = \gamma Y_{t-1,i} + \beta' X_{t,i} + U_{t,i}, \quad \text{with } (t,i) \in \mathbb{N}^2 \tag{1}$$

in which t denotes time at which quantities are measured and i denotes cross-sectional unit (individual).[1] The dependent variable $Y_{t,i} \in \mathbb{R}$ is assumed to depend on its own lag, as well as on a vector of time-varying explanatory variables $X_{t,i} \in \mathbb{R}^D$, all of which are observed. The disturbance or error term $U_{t,i} \in \mathbb{R}$ captures remaining unobserved individual heterogeneity. The standard setup is the one in which there is some small number T of time periods, while the number N of individuals in the panel is large. The usual asymptotic analysis is then as $N \to \infty$ while T remains fixed. It is worth pointing out at this stage that the results of our paper encompass the fixed T case as a special case; they are, however, derived in a more general setup in which both dimensions T and N get large (Hahn and Kuersteiner, 2002, 2004).

Let $\theta \equiv (\gamma, \beta')'$, with $|\gamma| < 1$ and $\beta \in \mathbb{R}^D$. The model in (1) is finitely parameterized, and the finite dimensional parameter θ is the focal point of the estimation problem. We treat the unknown (joint) distribution of the latent terms as an infinite dimensional nuisance parameter. Hence, our setup applies to a variety of panel models that are additive in the unobserved error.[2] Let $U_i \equiv (U_{1,i}, \dots, U_{T,i})' \in \mathbb{R}^T$, $X_i \equiv (X'_{1,i}, \dots, X'_{T,i})' \in \mathbb{R}^{DT}$ and $Y_i \equiv (Y_{1,i}, \dots, Y_{T,i})' \in \mathbb{R}^T$ be the vectors of individual i disturbances, explanatory and

[1] In what follows, we shall assume that the variables with non-positive time indices disappear from Equation (1). The main advantage of this assumption is that it simplifies the treatment of the initial observations. We note, however, that the latter is an important theoretical and practical problem in dynamic panel data models with unobserved errors (Arellano and Honoré, 2001; Wooldridge, 2005).

[2] In particular, by letting $U_{t,i} = \alpha_i + \varepsilon_{t,i}$ we can accommodate both linear and nonlinear panel models in which the effects α_i are random, provided the exogeneity conditions hold both for the errors $\varepsilon_{t,i}$ and the random effects α_i. Panel models with fixed effects can also be treated as part of our setup provided however

dependent variables observed throughout T. Assume that $\{(X_i', U_i')', i \geqslant 1\}$ is a sequence of vectors that are *independent and identically distributed* and consider estimating θ under the conditional moment restrictions that:

$$E(U_{t,i}|X_{t,i}, U_{t-1,i}, X_{t-1,i}, \ldots, U_{1,i}, X_{1,i}) = 0 \text{ for every } (t,i) \in [[1,T]] \times [[1,N]] \tag{2}$$

with probability one. Put in words, the conditions in (2) state that for any $(t,i) \in [[1,T]] \times [[1,N]]$, the error $U_{t,i}$ is mean independent of: (1) any of its predecessors $U_{h,i}$ with $h < t$, as well as (2) the contemporaneous and past values of the explanatory variable $X_{k,i}$, with $k \leqslant t$. The restrictions in Equation (2) say that the explanatory variables are *predetermined* (see e.g. Arellano and Honoré (2001)).

Because of the unknown (joint) distribution of the errors, the model defined by Equations (1) and (2) is semiparametric. This raises a question of semiparametric efficiency bound for the parameter of interest θ. Chamberlain (1992) is an important antecedent that provides a solution to this question; we now elaborate on Chamberlain's (1992) approach. Letting $Z_{t,i} \equiv (X_{t,i}', Y_{t,i})' \in \mathbb{R}^{D+1}$ be the vector of observables and $Q_{t,i} \equiv (X_{t,i}', U_{t-1,i})' \in \mathbb{R}^{D+1}$ be the vector of instruments both observed at time t for individual i, the orthogonality conditions in Equation (2) can be written in the form: $E[r_t(Z_i, \theta)|Q_{t,i}, \ldots, Q_{1,i}] = 0$, with probability one, for every $i = 1, \ldots, N$, where similar to previously we have defined $Z_i \equiv (Z_{1,i}', \ldots, Z_{T,i}') \in \mathbb{R}^{(D+1)T}$. Chamberlain (1992) derives the semiparametric efficiency bound for θ defined by such sequential conditional moment restrictions.

While few restrictions are placed on the form of the functions r_t, Chamberlain's (1992) approach crucially relies on the assumption that the vectors $Z_i, Q_{1,i}, \ldots, Q_{T,i}$ are iid across individuals. It is worth pointing out that this iid assumption does not restrict the heterogeneity and dependence structure of the variables across time. Indeed, different components of Z_i, for example, can be correlated—which results in an autocorrelated errors—and their marginal distributions need not be the same—which allows the errors to be heterogeneous across time. Still, the structure of the sequence $\{Z_i, i \geqslant 1\}$ across individuals is very rigid—independent

that they are linear, so that the standard approach of considering differences applies. One such example is Chamberlain (1992).

and identically distributed—and this property is crucial for Chamberlain's (1992) results to hold.

The reason behind lies in the very construction of the semiparametric efficiency bound. Similar to the setup used in Chamberlain (1987), Chamberlain (1992) relies on a multinomial approximation to the unknown distribution of Z_i. In the iid case, the efficient (in the sense of Hansen, 1982) GMM estimator of θ and the MLE obtained when the data is generated from a multinomial distribution are both asymptotically normally distributed with asymptotic covariance matrices respectively equal to Ω and I^{-1}, where I is the Fisher information matrix of the multinomial model. When the data has finite support, Chamberlain (1987) shows that Ω and I^{-1} are the same. Hence, they must be equal to the semiparametric efficiency bound for θ. Given that any distribution can be approximated arbitrarily well by a multinomial distribution, the general expression for the bound follows.

The iid assumption plays an important role in Chamberlain's (1987) construction of the semiparametric bounds. If the sequence $\{Z_i, i \geqslant 1\}$ is dependent and/or heterogeneous, Chamberlain's (1987) multinomial approximations no longer hold which makes the efficiency results in Chamberlain (1992) difficult to extend to dynamic panels with heterogeneous and cross section dependent unobserved effects.

Our approach to dealing with heterogeneity and cross section dependence is as follows. First, we adopt a framework that is more suited to both dimensions of the panel T and N getting large. We assume that random variables come in a form of a vector random field $\{Z_{t,s}, (t,s) \in \mathbb{N} \times \mathbb{S}\}$ indexed by a time index t and a space index s that takes values in some countable subset $\mathbb{S}$ of $\mathbb{R}^d$. For example, s could represent the coordinates of individuals placed on a grid in $\mathbb{R}^2$.[3] We allow the vectors $Z_{t,s}$ to be heterogeneous as well as weakly

[3]It is worth pointing out that we allow $Z_{t,s} \equiv (X'_{t,s}, U'_{t,s})' \in \mathbb{R}^{K+G}$ to be a vector of possibly large though fixed dimension; in particular, the "fixed T" panel setup may be obtained as a special case, by letting $G \equiv T$, $K \equiv DT$, and dropping the time index of the random field.

dependent across time and space. Specifically, we focus on the case in which the random field is strong (or α-) mixing.[4]

Since the variables in the random field are no longer independent and identically distributed, the order of indices matters for establishing the orthogonality conditions between the explanatory variables and disturbances. As a consequence, all of our results—including the expression of the semiparametric efficiency bound—change as one changes the ordering of t and s. While for the time index t, a unique order exists on $\mathbb{N}$, the same generally does not hold for the space index s (except for the case in which $\mathbb{S} \subseteq \mathbb{R}$). Our approach is to consider the order determined by an increasing sequence of subsets of $\mathbb{S}$: $\emptyset = \mathbb{S}_0 \subset \mathbb{S}_1 \subset \ldots \subset \mathbb{S}_{N-1} \subset \mathbb{S}_N \subset \ldots$ with cardinalities $|\mathbb{S}_N| = N$. Given such a sequence, we let $s_i \in \mathbb{S}_i \backslash \mathbb{S}_{i-1}$ for any $i \geqslant 1$, and impose the requirement that:

$$E(U_{t,s_i} | U_{t,s_j}, s_j \in \times \mathbb{S}_{i-1}, U_h, h \in [[1, t-1]], X) = 0 \tag{3}$$

with probability one, for any $(t, i) \in \mathbb{N}^2$, where we have let $U_h \equiv \{U_{h,s}, s \in \mathbb{S}\}$ and $X \equiv \{X_{t,s}, (t, s) \in \mathbb{N} \times \mathbb{S}\}$ be the entire fields of time-h errors, and explanatory variables, respectively. The property in (3) states that for any individual i and time t, U_{t,s_i} is mean independent of: (1) contemporaneous errors of its neighbors with space indices in S_{i-1}, (2) all past errors U_h ($h < t$), as well as (3) all explanatory variables. The conditional moment restrictions in (3) generalize the property of *strict exogeneity*—used in frameworks with iid variables—to heterogeneous dynamic panels with cross section dependence. Accordingly, we call this condition *panel exogeneity.*

Next, we show that the panel exogeneity condition is sufficient to characterize the semiparametric efficiency bound for the parameter of interest θ. For this, we translate the conditional moment conditions in Equation (3) into a moment condition on the observables $\{(X'_{t,s}, Y'_{t,s})', (t, s) \in \mathbb{N} \times \mathbb{S}\}$. Note that our panel exogeneity condition places no restriction on the joint behavior of X_{t,s_i} and X_{h,s_j}. In other words, the requirement in (3) allows the

[4]The notions of mixing for random fields are nontrivial extensions of their familiar counterparts used in time-series analysis. A detailed analysis is given in Ivanov and Leonenko (1986) and Föllmer (1988), for example.

explanatory variables to be heterogeneous and correlated across both time and individuals. This results in a very general setup, in which the dependent variable $Y_{t,s}$ is affected by its own lags, its own explanatory variables $X_{t,s}$, and—through the latter—the explanatory variables of other individuals in the panel.

The observable conditional moment restrictions are the starting point of our approach, which can be summarized as follows. Starting from the semiparametric model defined by the conditional moment restrictions on the observables, we consider fully parametric models that satisfy the same restriction, and contain the data generating process; these are called parametric *submodels* of the initial semiparametric model. Next, we look for one such parametric submodel that is the *least favorable*, in a sense that the inverse of the Fisher's information matrix I^{-1} is the largest. Whenever the latter equals the covariance matrix Ω of a feasible semiparametric estimator—such as GMM—it also equals the semiparametric efficiency bound. When no semiparametric estimator is available, then it still remains that the inverse Fisher information matrix provides a lower bound for the semiparametric efficiency.

The key insight behind our approach is due to Stein (1956). It is interesting to note that Chamberlain (1987, 1992) use similar argument. What is different, however, is the way of obtaining the least favorable submodel. As explained earlier, Chamberlain (1987) uses a multinomial approximation; our solution is to use the projection of the true but unknown conditional densities onto a set of densities that satisfy the conditional moment restriction. Because those densities are conditional, nothing in our projection approach restricts the variables of the random field to be iid.

Finally, we note that in time series models there are important results on the semiparametric efficiency with weakly dependent data. For example, Carrasco and Florens (2004) and Carrasco et al. (2007) compute the semiparametric efficiency bound for finite dimensional structural parameters known to satisfy a (potentially infinite) set of unconditional moment restrictions, though they focus on the time series that are strictly stationary.

The remainder of the paper is organized as follows: Section 2 describes our setup and we introduce the statistical models in Section 3. In Section 4 we characterize the least favorable parametric family. Section 5 derives the asymptotic distribution of the MLE in the latter, and concludes with the expression of the semiparametric efficiency bound. All proofs are relegated to Appendix.

2. Setup

2.1. **Structural Cross Section Time Series Model.** We shall consider models of spatial dependence among individuals. For this, let $Y_{t,s} \in \mathbb{R}^G$ ($G \geqslant 1$) denote the vector of dependent variables observed at time $t \in \mathbb{N}$, for an individual with spacial index $s \in \mathbb{S}$ where $\mathbb{S}$ is a countable subset of $\mathbb{R}^d$ ($d \geqslant 1$). For example, s can represent the geographical coordinates of the individual situated on a discrete grid in $\mathbb{R}^2$. Similarly, we shall denote $X_{t,s} \in \mathbb{R}^K$ ($K \geqslant 1$) the corresponding vector of explanatory variables, and $U_{t,s} \in \mathbb{R}^G$ the vector of unobserved errors.

Let an economic theory then specify the system of equations:

$$r(Y_{t,s}, \ldots, Y_{t-\tau,s}, X_{t,s}, \ldots, X_{t-\tau,s}, \theta) = U_{t,s}\ , \text{ for } (t,s) \in \mathbb{N} \times \mathbb{S} \tag{4}$$

The equations in (4) allow different components of $Y_{t,s}$ to depend not only on their own lags but also on different lags of the explanatory variables. The maximum number of lags τ is assumed known and fixed ($\tau \geqslant 0$). In what follows, we shall moreover assume that the functional form of r is known and that θ is finite dimensional so $\Theta \subset \mathbb{R}^k$.

We start by analyzing the system of equations in (4) when $(t,s) \in [[1,T]] \times \mathbb{S}_N$ where $T \geqslant 1$ and $\mathbb{S}_N$ forms an increasing sequence of subsets of $\mathbb{S}$, $\mathbb{S}_1 \subset \ldots \subset \mathbb{S}_N \subset \mathbb{S}_{N+1} \subset \ldots \subset \mathbb{S}$, with cardinality $|\mathbb{S}_N| = N$ ($N \geqslant 1$). This assumption is important as it will allow us to define an order of spacial indices in $\mathbb{S}$. While the ordering of time indices $t \in \mathbb{N}$ follows naturally from the observed direction of time, the same does not hold for the space indices $s \in \mathbb{S}$. It is worth pointing out that all of our results—including the expression for the semiparametric efficiency bound for the structural parameter θ in Equation (4)—depend on the way that both time and space indices t and s are ordered. This is in stark contrast from the earlier literature which assumes independence and identical distribution thus making the ordering of the indices irrelevant.

The setup in Equation (4) allows for the asymptotic analysis to be performed both as T is held fixed and $N \to \infty$, as well as when $(T, N) \to \infty$. The fixed T setup is obtained by letting $T = 1$ in Equation (4) and stacking different time lags of the variables into vectors $Y_{t,s}$, $X_{t,s}$ and $U_{t,s}$. In that case the dimension G of the vectors $Y_{t,s}$ and $U_{t,s}$ plays the traditional role of T. In what follows, we explicitly treat the second case $(T, N) \to \infty$ and derive the results corresponding to the fixed T benchmark as a special case.

We start by analyzing the system of equations in (4) when T and N are given. For this, we first group different vectors appearing in Equation (4) according to their time and space indices. Specifically, we let $\tilde{Y}_{TN} \equiv (Y_{t,s}, (t, s) \in [[1, T]] \times \mathbb{S}_N) \in \mathbb{R}^{GTN}$ be the collection of dependent variables observed up to time T for all space indices in $\mathbb{S}_N$. Similarly, we define $\tilde{X}_{TN} \in \mathbb{R}^{KTN}$ and $\tilde{U}_{TN} \in \mathbb{R}^{GTN}$, to be vectors of all the explanatory variables and unobserved errors, respectively.

The system of dynamic equations in (4) defines a mapping $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta) : \mathbb{R}^{GTN} \to \mathbb{R}^{GTN}$ from all the dependent variables $\tilde{Y}_{TN}$ to all the latent variables $\tilde{U}_{TN}$, and we write:

$$\tilde{r}(\tilde{Y}_{TN}, \tilde{X}_{TN}, \theta) = \tilde{U}_{TN}. \tag{5}$$

When $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ is differentiable at a point $y \in \mathbb{R}^{GTN}$, we shall denote by $J(y, \tilde{X}_{TN}, \theta)$ the Jacobian of $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ at y.[5] For any $y \in \mathbb{R}^{GTN}$ with components $y_{t,s} \in \mathbb{R}^G$ (and indices $(t, s) \in [[1, T]] \times \mathbb{S}_N$), we then have:

$$J(y, \tilde{X}_{TN}, \theta) = \prod_{t=1}^{T} \prod_{s \in \mathbb{S}_N} \det D_{Y_{t,s}} r_{t,s}(y_{t,s}, \theta), \tag{6}$$

where $r_{t,s}(Y_{t,s}, \theta) \equiv r(Y_{t,s}, \ldots, Y_{t-\tau,s}, X_{t,s}, \ldots, X_{t-\tau,s}, \theta)$ from Equation (4) and $D_{Y_{t,s}} r_{t,s}$ denotes the partial derivative of $r_{t,s}(Y_{t,s}, \theta)$ with respect to its first variable $Y_{t,s}$.

[5]Hereafter, by derivative of $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ we mean a linear transformation $D_{\tilde{Y}_{TN}} \tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ of $\mathbb{R}^{GTN}$ into $\mathbb{R}^{GTN}$—property which we denote $D_{\tilde{Y}_{TN}} \tilde{r}(\cdot, \tilde{X}_{TN}, \theta) \in L(\mathbb{R}^{GTN}, \mathbb{R}^{GTN})$—such that for any $(y, h) \in \mathbb{R}^{2GTN}$ we have:

$$\lim_{h \to 0} \frac{|\tilde{r}(y + h, \tilde{X}_{TN}, \theta) - \tilde{r}(y, \tilde{X}_{TN}, \theta) - D_{\tilde{Y}_{TN}} \tilde{r}(y, \tilde{X}_{TN}, \theta)' h|}{|h|} = 0.$$

The Jacobian $J(y, \tilde{X}_{TN}, \theta)$ is then given by: $J(y, \tilde{X}_{TN}, \theta) = \det D_{\tilde{Y}_{TN}} \tilde{r}(y, \tilde{X}_{TN}, \theta)$.

We now provide conditions under which—given the explanatory variables and the parameter θ—the system in Equation (5) defines a homeomorphic mapping from the the latent variables $\tilde{U}_{TN}$—the unobservables—to the dependent variables $\tilde{Y}_{TN}$—the observables.[6] We shall assume:

Assumption A1. For any $T \geqslant 1$, $N \geqslant 1$, every $(\tilde{X}_{TN}, \theta) \in \mathbb{R}^{KTN} \times \Theta$ and $(t, s) \in [[1, T]] \times \mathbb{S}_N$: (i) the map $(y_1, \dots, y_{\tau+1}) \mapsto r(y_1, \dots, y_{\tau+1}, X_{t,s}, \dots, X_{t-\tau,s}, \theta)$ is in $\mathcal{C}^1(\mathbb{R}^{G(\tau+1)}, \mathbb{R}^G)$; (ii) $\det D_{Y_{t,s}} r_{t,s}$ never vanishes; (iii) $\lim_{|(y'_1,\dots,y'_{\tau+1})'| \to +\infty} |r(y_1, \dots, y_{\tau+1}, X_{t,s}, \dots, X_{t-\tau,s}, \theta)| = +\infty$.

Assumption A1(i) implies that the mapping $\tilde{r}$ in Equation (5) is continuously differentiable with respect to the dependent variables $\tilde{Y}_{TN}$. Moreover, using the equality established in (6) together with Assumption A1(ii) we know that $J(y, \tilde{X}_{TN}, \theta) \neq 0$ for any $y \in \mathbb{R}^{GTN}$. These two conditions on $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ are sufficient to apply the Implicit Function Theorem: given the explanatory variables and the parameter θ, Equation (5) can be solved for the dependent variables in terms of the latent variables in a neighborhood of any point $(u, y) \in \mathbb{R}^{2GTN}$ at which $\tilde{r}(y, \tilde{X}_{TN}, \theta) = u$ and $J(y, \tilde{X}_{TN}, \theta) \neq 0$ (see e.g. Theorem 9.28 in Rudin (1976)).

The resulting mapping is a *local* homeomorphism: it maps an open neighborhood of u homeomorphically onto an open neighborhood of y. This condition by itself does not insure that this mapping is either one-to-one or onto.[7] It is by adding the condition in Assumption A1(iii) that we can guarantee that the mapping from the latent variables $\tilde{U}_{TN}$ to the dependent variables $\tilde{Y}_{TN}$ is also a homeomorphism of $\mathbb{R}^{GTN}$ onto itself. Assumption A1(iii) implies that $|\tilde{r}(y, \tilde{X}_{TN}, \theta)|$ goes to infinity as $|y|$ gets large ($y \in \mathbb{R}^{GTN}$); this is a necessary and sufficient condition that the mapping $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ be proper, i.e. that the inverse image by $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ of any compact set in $\mathbb{R}^{GTN}$ be a compact in $\mathbb{R}^{GTN}$. We then obtain the following result:

Proposition 1. *Let the system of dynamic equations be defined as in Equation* (5)*, and let Assumption A1 hold. Then, given any $T \geqslant 1$, $N \geqslant 1$, and $(\tilde{X}_{TN}, \theta) \in \mathbb{R}^{KTN} \times \Theta$, the*

[6] A definition of homeomorphic is: continuous, one-to-one, onto, and having a continuous inverse.

[7] Standard counterexample is the mapping $(x_1, x_2) \to (\exp x_1 \cos x_2, \exp x_1 \sin x_2)$.

transformation from $\tilde{U}_{TN}$ *to* $\tilde{Y}_{TN}$ *is a diffeomorphism of* $\mathbb{R}^{GTN}$ *onto itself, and we denote it* $\tilde{q}(\cdot, \tilde{X}_{TN}, \theta) : \mathbb{R}^{GTN} \to \mathbb{R}^{GTN}$.

In other words, the transformation $\tilde{q}(\cdot, \tilde{X}_{TN}, \theta)$ from the unobservables to the observables is differentiable, one-to-one in $\mathbb{R}^{GTN}$ onto itself, and has a differentiable inverse. Obviously, $\tilde{q}(\cdot, \tilde{X}_{TN}, \theta)$ is then homeomorphic.

Proposition 1 implies the following: if we give ourselves a set of explanatory variables $\tilde{X}_{TN}$ and a set of latent variables $\tilde{U}_{TN}$, then the dynamic system in Equation (5) allows us to determine the dependent variables $\tilde{Y}_{TN}$ for any given value of the parameter θ through a set of reduced form equations: $\tilde{Y}_{TN} = \tilde{q}(\tilde{U}_{TN}, \tilde{X}_{TN}, \theta)$. The dynamic system in Equation (5) is therefore *complete* because it accounts for the formation of the values of *all* the dependent variables. We now turn to a probabilistic description of how the variables in Equation (5) are generated.

2.2. **Random Fields.** For any $(t, s) \in \mathbb{N} \times \mathbb{S}$, define:

$$W_{t,s} \equiv (X'_{t,s}, U'_{t,s})' \in \mathbb{R}^{K+G}, \tag{7}$$

and consider $W \equiv \{W_{t,s}, (t, s) \in \mathbb{N}\times\mathbb{S}\}$ to be a collection of random vectors—a vector random field—defined on a probability space $(\Omega, \mathcal{W}, P)$ where $W : \Omega \to (\mathbb{R}^{K+G})^{\mathbb{N}\times\mathbb{S}}$, and $(\mathbb{R}^{K+G})^{\mathbb{N}\times\mathbb{S}}$ is the product space generated by taking a copy of $\mathbb{R}^{K+G}$ for each element in $\mathbb{N} \times \mathbb{S}$. Given $T \geqslant 1$ and $N \geqslant 1$, we focus on the TN components of W indexed by $(t, s) \in [[1, T]] \times \mathbb{S}_N$, which we denote $\tilde{W}_{TN} \in \mathbb{R}^{(K+G)TN}$. We let $\tilde{\mathcal{W}}_{TN} \equiv \sigma(W_{t,s}, (t, s) \in [[1, T]] \times \mathbb{S}_N)$. Note that the σ-algebra $\mathcal{W}$ contains all the information generated by the random vectors $W_{t,s}$, i.e. $\tilde{\mathcal{W}}_{TN} \subset \mathcal{W}$ for every $T \geqslant 1$ and $N \geqslant 1$.

Let $\tilde{\mathcal{X}}_{TN}$ be a sub-σ-field of $\mathcal{W}_{TN}$ generated by the explanatory variables alone: $\tilde{\mathcal{X}}_{TN} \equiv \sigma(X_{t,s}, (t, s) \in [[1, T]]\times\mathbb{S}_N)$. We denote by μ_{TN} a regular conditional probability distribution (measure) for $\tilde{U}_{TN}$ given $\tilde{\mathcal{X}}_{TN}$, i.e. $\mu_{TN} : \Omega \times \mathcal{B}^{GTN} \to \mathbb{R}_+$ satisfies: (i) for each $B \in \mathcal{B}^{GTN}$,

$\omega \mapsto \mu_{TN}(\omega, B)$ is a version of $P(\tilde{U}_{TN}(\omega) \in B|\tilde{\mathcal{X}}_{TN})$, and (ii) for a.e. ω, $B \mapsto \mu_{TN}(\omega, B)$ is a probability measure on $\{\mathbb{R}^{GTN}, \mathcal{B}^N\}$.[8] As usual, $\mathcal{B}^{GTN}$ denotes the Borel σ-field on $\mathbb{R}^{GTN}$.

To simplify, we assume that our economic theory further specifies that a regular conditional distribution for $\tilde{U}_{TN}$ given $\tilde{\mathcal{X}}_{TN}$ is absolutely continuous (with respect to Lebesgue measure).[9] So by Radon-Nikodym theorem, for a.e. ω:

$$\mu_{TN}(\omega, B) = \int_B f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}(u)(\omega)du,$$

where $f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}} : \Omega \times \mathbb{R}^{GTN} \to \mathbb{R}_+$, and $f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}(\cdot)(\omega)$ is a conditional density of $\tilde{U}_{TN}$ given $\tilde{\mathcal{X}}_{TN}$. Note that both μ_{TN} and $f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}$ are random elements, hence any statements made about them are to be understood to hold P almost surely (a.s.).

Now, recall that for any given $(\tilde{X}_{TN}, \theta) \in \mathbb{R}^{KTN} \times \Theta$, the mapping $q(\cdot, \tilde{X}_{TN}, \theta)$ defined in Proposition 1 is diffeomorphic from $\mathbb{R}^{GTN}$ to $\mathbb{R}^{GTN}$. This has two important implications. First, by letting $\nu_{TN} : \Omega \times \mathcal{B}^{GTN} \to \mathbb{R}_+$ be defined as:

$$\nu_{TN}(\omega, A) \equiv \mu_{TN}(\omega, q^{-1}(A, \tilde{X}_{TN}(\omega), \theta)), \tag{8}$$

we have that for a.e. ω, $A \mapsto \nu_{TN}(\omega, A)$ is a measure on $\{\mathbb{R}^{GTN}, \mathcal{B}^N\}$; and for each $A \in \mathcal{B}^{GTN}$, $\omega \mapsto \nu_{TN}(\omega, A)$ is a version of $P(\tilde{U}_{TN}(\omega) \in q^{-1}(A, \tilde{X}_{TN}(\omega), \theta)|\tilde{\mathcal{X}}_{TN}) = P(\tilde{Y}_{TN}(\omega) \in A|\tilde{\mathcal{X}}_{TN})$ (see e.g. Theorem 3.21 in Davidson (1994)). In other words, ν_{TN} is a regular conditional distribution (measure) for $\tilde{Y}_{TN}$ given $\tilde{\mathcal{X}}_{TN}$. Since $q(\cdot, \tilde{X}_{TN}, \theta)$ is a homeomorphism, it is equivalent to work with either μ_{TN} or ν_{TN}: we can go from one measure to another by the mapping $\tilde{r}(\cdot, \tilde{X}_{TN}, \theta)$ or its inverse $q(\cdot, \tilde{X}_{TN}, \theta)$. Second, the measure ν_{TN} is absolutely

---

[8]Regular conditional distributions allow us to simultaneously compute the conditional expectations of all functions of $\tilde{U}_{TN}$. For a proof of their existence see e.g. Theorem 4.1.6 in Durrett (1995).

[9]A common example is the one in which the probability measure of $\tilde{W}_{TN}$ is absolutely continuous (with respect to Lebesgue measure) with density $f_{\tilde{W}_{TN}}$ that is positive on $\mathbb{R}^{(K+G)TN}$. Then, for every $(x, B) \in \mathbb{R}^{KTN} \times \mathcal{B}^{GTN}$, we can let:

$$Q(x, B) \equiv \int_B f_{\tilde{W}_{TN}}(u, x)du \Big/ \int_{\mathbb{R}^{GTN}} f_{\tilde{W}_{TN}}(u, x)du,$$

and $\mu_{TN}(\omega, B) \equiv Q(\tilde{X}_{TN}(\omega), B)$ is a regular conditional distribution for $\tilde{U}_{TN}$ given $\tilde{\mathcal{X}}_{TN}$ that is absolutely continuous (with respect to Lebesgue measure) continuous (with respect to Lebesgue measure) with density $f_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}$ given by:

$$f_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(y)(\omega) = f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{r}(y, \tilde{X}_{TN}(\omega), \theta))(\omega)|J(y, \tilde{X}_{TN}(\omega), \theta)| \text{ a.s.,} \tag{9}$$

for any $y \in \mathbb{R}^{GTN}$ where $J(y, \tilde{X}_{TN}(\omega), \theta)$ is the Jacobian of $\tilde{r}(\cdot, \tilde{X}_{TN}(\omega), \theta)$ at y, defined previously (see e.g. Theorem 8.18 in Davidson (1994)). Hereafter, we shall drop reference to ω in the expressions involving conditional densities—such as the one in Equation (9)—whenever doing so does not introduce any ambiguity.

In what follows, we shall assume the following:

Assumption A2. For any $T \geqslant 1$ and $N \geqslant 1$, the joint distribution of $\tilde{X}_{TN}$ does not depend on θ.

To give an insight into Assumption A2, consider a sample $(\tilde{x}'_{TN}, \tilde{y}'_{TN})' \in \mathbb{R}^{(K+G)TN}$ of observations of the explanatory and dependent variables $\tilde{X}_{TN}$ and $\tilde{Y}_{TN}$.[10] Then, the statement in A2 implies that all the sample information concerning the parameter of interest θ can be obtained from the partial likelihood function $f_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{y}_{TN}) = f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{r}(\tilde{y}_{TN}, \tilde{x}_{TN}, \theta))|J(\tilde{y}_{TN}, \tilde{x}_{TN}, \theta)|$, and the distribution of $\tilde{X}_{TN}$ need not even be specified. This separability property of the likelihood function is discussed by Engle et al. (1983), for example.

We now construct statistical models for the joint conditional densities $f_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}$.

3. Statistical Models

3.1. **Exogeneity Condition.**

The key tool used in the construction of our statistical models for $f_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}$ is a set of exogeneity conditions between $\tilde{U}_{TN}$ and $\tilde{X}_{TN}$, to which we now turn.

Note that $f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}$ in Equation (9) is a joint conditional density of $\tilde{U}_{TN}$ given $\tilde{\mathcal{X}}_{TN}$. We can further "decompose" $f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}$ into a sequence of marginal conditional densities. For this, let $U_t \equiv (U_{t,s}, s \in \mathbb{S}_N) \in \mathbb{R}^{GN}$ and $Y_t \equiv (Y_{t,s}, s \in \mathbb{S}_N) \in \mathbb{R}^{GN}$ be the vectors of disturbances and dependent variables, respectively, observed at times $t \in [[1, T]]$ for all individuals with space indices in $\mathbb{S}_N$. Then let $\mathcal{G}_t \equiv \sigma(U_h, h \in [[1, t-1]], \tilde{X}_{TN})$, and $\mathcal{F}_t \equiv \sigma(Y_h, h \in [[1, t-1]], \tilde{X}_{TN})$,

[10]Realizations of random variables (or vectors), e.g. $X_{t,s}$, are denoted using lowercase letters, e.g. $x_{t,s}$.

for any $t \in [[1,T]]$.[11] For any $(u,y) \in \mathbb{R}^{2GN}$ with components $(u_t, y_t), t \in [[1,T]]$, we then have:

$$f_{\tilde{U}_{TN}|\tilde{\mathcal{X}}_{TN}}(u) = \prod_{t=1}^{T} f_{U_t|\mathcal{G}_t}(u_t) \quad \text{and} \quad f_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(y) = \prod_{t=1}^{T} f_{Y_t|\mathcal{F}_t}(y_t) \text{ a.s..} \tag{10}$$

The densities $f_{U_t|\mathcal{G}_t}$ and $f_{Y_t|\mathcal{F}_t}$ are still joint conditional densities of all random vectors $U_{t,s}$ and $Y_{t,s}$, respectively, as s takes values in $\mathbb{S}_N$. We now decompose them in a sequence of conditional densities, where the conditioning is done with respect to the spacial indices s. We start by defining the conditioning sets. For this, let the elements of $\mathbb{S}_N$ be labeled as follows: $s_1 \in \mathbb{S}_1$, $s_2 \in \mathbb{S}_2 \backslash \mathbb{S}_1$, up to $s_N \in \mathbb{S}_N \backslash \mathbb{S}_{N-1}$. Obviously, the labeling of elements in $\mathbb{S}_N$ depends on the way the sequence of sets $\mathbb{S}_1, \dots, \mathbb{S}_{N-1}$ increases up to $\mathbb{S}_N$. Now, for any $i \in [[1,N]]$, let $\mathcal{G}_{t,s_i} \equiv \sigma(U_{t,v}, v \in \mathbb{S}_{i-1}, U_h, h \in [[1,t-1]], \tilde{X}_{TN})$, and $\mathcal{F}_{t,s_i} \equiv \sigma(Y_{t,v}, v \in \mathbb{S}_{i-1}, Y_h, h \in [[1,t-1]], \tilde{X}_{TN})$.[12] Then, for any $T \geqslant 1$, every $t \in [[1,T]]$, and any $N \geqslant 1$, we can write:

$$f_{U_t|\mathcal{G}_t}(u_t) = \prod_{i=1}^{N} f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}(u_{t,s_i}) \quad \text{and} \quad f_{Y_t|\mathcal{F}_t}(y_t) = \prod_{i=1}^{N} f_{Y_{t,s_i}|\mathcal{F}_{t,s_i}}(y_{t,s_i}) \text{ a.s..} \tag{11}$$

As pointed out previously, the labeling of spacial indices in Equation (11) depends on the way the sequence of space index sets $\mathbb{S}_1, \dots, \mathbb{S}_{N-1}$ increases up to $\mathbb{S}_N$. We are now ready to state the *panel exogeneity* condition used throughout this paper:

Assumption A3. For any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, the conditional distribution of U_{t,s_i} given $\mathcal{G}_{t,s_i}$ satisfies: $E(U_{t,s_i}|\mathcal{G}_{t,s_i}) \equiv \int_{\mathbb{R}^G} u f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}(u) du = 0$ a.s..

The above definition generalizes the notion of strict exogeneity employed in systems in which the variables are independent and identically distributed across individuals (see e.g. Arellano and Honoré (2001)) to dynamic systems with heterogeneity and cross section dependence. Under Assumption A3, for any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$,

[11] For $t = 1$, we let $\mathcal{G}_1 \equiv \tilde{\mathcal{X}}_{TN}$ and $\mathcal{F}_1 \equiv \tilde{\mathcal{X}}_{TN}$.

[12] For $i = 1$, we let $\mathcal{G}_{t,s_1} \equiv \mathcal{G}_t$ and $\mathcal{F}_{t,s_1} \equiv \tilde{\mathcal{F}}_t$.

the unobserved error U_{t,s_i} is mean independent of: (1) any set of contemporaneous neighbors' errors U_{t,s_j} with $j < i$, (2) all the lags U_h with $h < t$, as well as (3) all the explanatory variables $\tilde{X}_{TN}$ (and all measurable functions thereof).

It is worth pointing out that the conditional moment restrictions in A3 incorporate the fact that U_{t,s_i} is conditionally mean independent of any future values of the explanatory variables X_{k,s_i} with $k > t$. This condition is equivalent to saying that $E[U_{t,s_i}\psi(X_{k,s_i})] = 0$ for any measurable function ψ. The latter is however weaker than requiring that for any measurable χ we also have $E[\chi(U_{t,s_i})\psi(X_{k,s_i})] = 0$, which is itself equivalent to independence of U_{t,s_i} and X_{k,s_i} $(k > t)$. In other words, the panel exogeneity condition in Assumption A3 allows for the possibility that current values of $X_{t,s}$ be influenced by past values of the disturbances.

The set of conditional expectation restrictions in Assumption A3 translates into a restriction on the observables. Combining Equations (6), (9), (10) and (11), and using the fact that they hold for any choice of $T \geqslant 1$ and $N \geqslant 1$, we have:

$$f_{Y_{t,s_i}|\mathcal{F}_{t,s_i}}(y_{t,s_i}) = f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}(r_{t,s_i}(y_{t,s_i}, \theta))|\det D_{Y_{t,s_i}} r_{t,s_i}(y_{t,s_i}, \theta)|, \text{ a.s.} \tag{12}$$

for every $t \in [[1, T]]$ and every $i \in [[1, N]]$. Now, take any $(t, i) \in [[1, T]] \times [[1, N]]$ in Equation (12) and consider the moment condition in Assumption A3. A simple change of variable $u = r_{t,s_i}(y, \theta)$ then yields:

$$\begin{aligned}
\int_{\mathbb{R}^G} u f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}(u)du &= \int_{\mathbb{R}^G} r_{t,s_i}(y, \theta) f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}(r_{t,s_i}(y, \theta))|\det D_{Y_{t,s_i}} r_{t,s_i}(y, \theta)|dy \\
&= \int_{\mathbb{R}^G} r_{t,s_i}(y, \theta) f_{Y_{t,s_i}|\mathcal{F}_{t,s_i}}(y)dy \\
&\equiv E[r_{t,s_i}(Y_{t,s_i}, \theta)|\mathcal{F}_{t,s_i}] = 0 \text{ a.s., for every } (t, i) \in [[1, T]] \times [[1, N]], \qquad (13)
\end{aligned}$$

where the first equality uses the fact that for a given value of θ, the mapping from U_{t,s_i} to Y_{t,s_i} is diffeomorphic from $\mathbb{R}^G$ onto itself as implied by Proposition 1; the second equality follows by Equation (11) (for a change of variables result see e.g. Theorem 10.9 in Rudin (1976)).

Put in words, restricting the conditional expectation of U_{t,s_i} under $f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}$ is equivalent to restricting the conditional expectation of $r_{t,s_i}(Y_{t,s_i},\theta)$ under $f_{Y_{t,s_i}|\mathcal{F}_{t,s_i}}$. The set of conditional moment restrictions in Equation (13) is the basis of our semiparametric model, to which we turn next.

3.2. **Semiparametric Model.** In what follows we shall assume that the endogenous variables are generated from the dynamic system in (4), in which $\theta = \theta_0$ and the conditional distribution of the latent variables has the following property:

Assumption A4. For any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, we have: $f_{U_{t,s_i}|\mathcal{G}_{t,s_i}}(u) > 0$ a.s. for every $u \in \mathbb{R}^G$.

Fixing any $(t,i) \in [[1,T]] \times [[1,N]]$ in Equation (12) and using the property in Assumption A4, then shows that the conditional distribution of the endogenous variable Y_{t,s_i} given $\mathcal{F}_{t,s_i}$ needs to be positive on $\mathbb{R}^G$ a.s.. To simplify the notation, we shall hereafter denote the latter as f_{t,s_i}, i.e. for every $y \in \mathbb{R}^G$ we let $f_{t,s_i}(y) \equiv f_{Y_{t,s_i}|\mathcal{F}_{t,s_i}}(y)$.

Our *semiparametric* model for f_{t,s_i}, denoted $\mathcal{SP}_{t,s_i}$, is defined as a set of all positive conditional densities on $\mathbb{R}^G$ that are measurable with respect to the information set $\mathcal{F}_{t,s_i}$ and satisfy the conditional moment restrictions in Equation (13).[13] More formally, we let $\mathcal{SP} \equiv \bigcup_{\theta\in\Theta} \mathcal{SP}_\theta$ with $\mathcal{SP}_\theta \equiv \bigotimes_{t\in[[1,T]]} \bigotimes_{i\in[[1,N]]} \mathcal{SP}_{\theta,t,s_i}$, and $\mathcal{SP}_{\theta,t,s_i}$ defined as:

$$\begin{aligned} \mathcal{SP}_{\theta,t,s_i} \equiv \Big\{ & g : \Omega \times \mathbb{R}^G \to \mathbb{R}_+ \text{ s.t.: (i) for a.e. } \omega,\ g(y)(\omega) > 0 \text{ for every } y \in \mathbb{R}^G \text{ and} \\ & \int_{\mathbb{R}^G} g(y)(\omega)dy = 1;\ \text{(ii) for every } y \in \mathbb{R}^G,\ g(y)(\omega) \text{ is } \mathcal{F}_{t,s_i}\text{-measureable; and (iii)} \\ & \int_{\mathbb{R}^G} r_{t,s_i}(y,\theta) g(y)(\omega)dy = 0 \text{ for a.e. } \omega \Big\} \end{aligned} \tag{14}$$

The model $\mathcal{SP}_{\theta,t,s_i}$ defined by Equation (14) is a collection of conditional probability densities g given $\mathcal{F}_{t,s_i}$ that are positive on $\mathbb{R}^G$. As a such, $\mathcal{SP}_{\theta,t,s_i}$ is parameterized by an infinite dimensional parameter ranging over an infinite dimensional space that describes the probability densities. Those densities are however restricted in an important way: the

[13] The idea of defining a statistical model as a collection of probability densities—in the context of dynamic systems such as the one in Equation (5)—is traceable to Koopmans (1950).

conditional expectations of $r_{t,s_i}(Y_{t,s_i}, \theta)$ under g are equal to zero. Hence, each model $\mathcal{SP}_{\theta,t,s_i}$ is also parameterized by a parametric component θ ranging over a finite dimensional space Θ. This gives the model $\mathcal{SP}$ its semiparametric structure.

Equation (13) shows that under Assumption A3 our semiparametric model $\mathcal{SP}$ is correctly specified: for any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, we have $f_{t,s_i} \in \mathcal{SP}_{\theta_0,t,s_i}$ where we have defined θ_0 to be the true value of the structural parameter θ in Equation (4). More conditions are needed to ensure that $\mathcal{SP}$ is also (point) identified, i.e. that $f_{t,s_i} \in \mathcal{SP}_{\theta_0,t,s_i}$ implies that for any $\theta_1 \in \Theta \backslash \{\theta_0\}$ we have $f_{t,s_i} \notin \mathcal{SP}_{\theta_1,t,s_i}$. For this we impose the following:

Assumption A5. For any $T \geqslant 1$, $N \geqslant 1$, every $(t,i) \in [[1,T]] \times [[1,N]]$, and every $(\theta_1, \theta_2) \in \Theta^2$, we have: $E[r_{t,s_i}(Y_{t,s_i}, \theta_1)|\mathcal{F}_{t,s_i}] = E[r_{t,s_i}(Y_{t,s_i}, \theta_2)|\mathcal{F}_{t,s_i}]$ a.s. only if $\theta_1 = \theta_2$.

Note that Assumption A5 imposes restrictions on a global behavior of the structural function r in Equation (4). It is therefore stronger than requiring that any parameter value θ_1 be locally identified in Θ. Local identification could for example be achieved by imposing conditions on $\det D_{Y_{t,s_i}} r_{t,s_i}(Y_{t,s_i}, \theta_1)$ which would insure that Implicit Function Theorem applies in a neighborhood of any θ_1 at which $E[r_{t,s_i}(Y_{t,s_i}, \theta_1)|\mathcal{F}_{t,s_i}] = 0$ a.s..

3.3. **Parametric Submodel.** In general, the model $\mathcal{SP}_{\theta,t,s_i}$ in Equation (14) contains more than one element: this means that the structural parameter θ generally does not characterize one conditional probability distribution. In other words, $\mathcal{SP}_{\theta,t,s_i}$ may contain several parametric families of known distributions $f_{\theta,\pi}$ possibly indexed by an additional finite dimensional parameter $\pi \in \Pi$, $\Pi \subset \mathbb{R}^p$ with $p < \infty$. When one such parametric family is in addition correctly specified, we call it a *parametric submodel* of $\mathcal{SP}$:

Definition 1. $\mathcal{P} \equiv \bigcup_{\theta\in\Theta} \bigcup_{\pi\in\Pi} \mathcal{P}_{\theta,\pi}$ with $\mathcal{P}_{\theta,\pi} \equiv \bigotimes_{t\in[[1,T]]} \bigotimes_{i\in[[1,N]]} \mathcal{P}_{\theta,\pi,t,s_i}$, and $\mathcal{P}_{\theta,\pi,t,s_i} \equiv \{\, f_{\theta,\pi} : \Omega \times \mathbb{R}^G \to \mathbb{R}_+$ s.t.: (i) for a.e. ω, $f_{\theta,\pi}(y)(\omega) > 0$ for every $y \in \mathbb{R}^G$ and $\int_{\mathbb{R}^G} f_{\theta,\pi}(y)(\omega)dy = 1$; (ii) for every $y \in \mathbb{R}^G$, $f_{\theta,\pi}(y)(\omega)$ is $\mathcal{F}_{t,s_i}$-measureable$\}$, is a parametric submodel of $\mathcal{SP}$ if: (i) for every $(\theta,\pi) \in \Theta \times \Pi$ we have $\mathcal{P}_{\theta,\pi,t,s_i} \subset \mathcal{SP}_{\theta,t,s_i}$, and (ii) for some $\pi_0 \in \Pi$ we have $f_{t,s_i} \in \mathcal{P}_{\theta_0,\pi_0,t,s_i}$.

The idea of such parametric submodels is traceable back to Stein (1956). Each $\mathcal{P}_{\theta,\pi,t,s_i}$ is a *parametric* model for f_{t,s_i} since every element $f_{\theta,\eta}$ of $\mathcal{P}_{\theta,\pi,t,s_i}$ is parameterized by a finite dimensional parameter $(\theta,\pi) \in \Theta \times \Pi$. Within each parametric submodel $\mathcal{P}$, the true value θ_0 of the structural parameter can be estimated via maximum likelihood. Let $\Omega_{\mathcal{P},\theta_0}$ denote its asymptotic covariance matrix. Since the semiparametric estimation of θ is always at least as difficult as the fully parametric one (Stein, 1956; Bickel, 1982), the *semiparametric efficiency bound* V_{θ_0} for θ_0 is larger than the supremum of $\Omega_{\mathcal{P},\theta_0}$ over the parametric submodels $\mathcal{P}$. If for some $\mathcal{P}^*$ this supremum is attained, then $\mathcal{P}^*$ is called *least favorable* parametric submodel: in that case $\Omega_{\mathcal{P}^*,\theta_0} = \sup_{\mathcal{P}\subset\mathcal{SP}} \Omega_{\mathcal{P},\theta_0}$.

4. Least Favorable Family

In this section we show that the least favorable parametric submodel $\mathcal{P}^*$ of the semiparametric model $\mathcal{SP}$ can be constructed by projecting the true but unknown densities f_{t,s_i} onto the sets $\mathcal{SP}_{\theta,t,s_i}$. Since for any given $(t,i) \in [[1,T]] \times [[1,N]]$ with $T \geqslant 1$ and $N \geqslant 1$ fixed, the projection of f_{t,s_i} onto $\mathcal{SP}_{\theta,t,s_i}$ is unique—up to an unavoidable P equivalence—the submodel $\mathcal{P}^*_{\theta,\pi,t,s_i}$ consists of a single element which we shall denote $g^*_{t,s_i}(\cdot,\theta)$.[14] In other words, the least favorable density $g^*_{t,s_i}(\cdot,\theta)$ is parameterized by the structural parameter θ alone. This approach is particularly appealing because it naturally leads to a maximum likelihood estimator for the true value θ_0 of θ, and—following Stein's (1956) idea—to its semiparametric efficiency bound.

4.1. **Projection Approach.**

The fundamental building block of the probability densities projections we are interested in is a class of distances defined on sets of probability densities. Specifically, a distance $\mathcal{D}$ on a set S is any nonnegative valued function defined on $S \times S$ such that $\mathcal{D}(s,m) = 0$ if and only if $s = m$. For $S_0 \subset S$, we write: $\mathcal{D}(S_0,m) = \inf_{s\in S_0} \mathcal{D}(s,m)$. To make the projection operational we need to define the distances upon which the $\mathcal{D}$-projection

[14]We shall say that two elements f and g of $\mathcal{SP}_{\theta,t,s_i}$ are P equivalent (or belong to the same P equivalence class) if and only if $f = g$ a.s..

are based. Given a convex function ϕ, a ϕ-divergence between s and m, denoted $\mathcal{D}_\phi(s,m)$, is formally defined as $\mathcal{D}_\phi(s,m) \equiv \int m(y)\phi\left(s(y)/m(y)\right)dy$ (Ali and Silvey, 1966; Csiszar, 1967).

Here, we are interested in the $\mathcal{D}_\phi$-projection of f_{t,s_i} onto a set $\mathcal{SP}_{\theta,t,i}$ defined as follows:

Definition 2. For any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, the $\mathcal{D}_\phi$-projection of f_{t,s_i} onto a set $\mathcal{SP}_{\theta,t,s_i}$ is (when it exists) a density $g^*_{t,s_i}(\cdot,\theta) \in \mathcal{SP}_{\theta,t,s_i}$ satisfying $\mathcal{D}_\phi(g^*_{t,s_i}(\cdot,\theta), f_{t,s_i}) = \mathcal{D}_\phi(\mathcal{SP}_{\theta,t,s_i}, f_{t,s_i})$ a.s., where for any $g \in \mathcal{SP}_{\theta,t,s_i}$ we let:

$$\mathcal{D}_\phi(g, f_{t,s_i}) \equiv E\left[\phi\left(\frac{g(Y_{t,s_i})}{f_{t,s_i}(Y_{t,s_i})}\right)\middle|\mathcal{F}_{t,s_i}\right] = \int_{\mathbb{R}^G}\phi\Big(\frac{g(y)}{f_{t,s_i}(y)}\Big)f_{t,s_i}(y)dy. \tag{15}$$

We restrict the class of divergences $\mathcal{D}_\phi$ in Equation (15) by considering the functions ϕ with the following properties:

Assumption A6. (i) $\phi \in \mathcal{C}^4(\mathbb{R}^*_+, \mathbb{R}_+)$; (ii) ϕ is strictly convex; (iii) $\phi(1) = \phi'(1) = 0$, $\phi''(1) = 1$; (iv) $\lim_{u\to+\infty}\phi'(u) = +\infty$; (v) $\phi'(0) = \lim_{u\to 0}\phi'(u) = -\infty$.[15]

Assumptions A6(i)-(iii) are fairly standard. Notice that requirement A6(iv) rules out the reverse I-divergence, $\phi(u) = -\ln u + u - 1$, and the Hellinger distance, $\phi(u) = (\sqrt{u}-1)^2$, since for both cases $\lim_{u\to+\infty}\phi'(u) = 0$. Assumption A6(v) is not strictly necessary, but it guarantees that the $\mathcal{D}_\phi$-projection of f_{t,s_i} is almost everywhere positive, provided such density is in $\mathcal{SP}_{\theta,t,s_i}$.

An important example of ϕ satisfying all the requirements of Assumption A6 is $\phi(u) = u\ln u - u + 1$; then the distance in Equation (17) is equivalent to the Kullback-Leibler information criterion (KLIC) or I-divergence (see e.g. Kullback and Khairat (1966) and Csiszar (1975) for a detailed study of the corresponding $\mathcal{D}_\phi$-projection). In the econometric literature, an application of I-divergence can be found in Kitamura and Stutzer (1997)'s Exponential Tilting estimator.

[15]Further, to deal with zeros, we adopt the understanding that:

$$\phi(0) = \lim\nolimits_{u\to0}\phi(u), \quad \phi'(0) = \lim\nolimits_{u\to0}\phi'(u), \quad 0\cdot\phi\left(\frac{v}{0}\right) = v\cdot\lim\nolimits_{u\to+\infty}\phi'(u) = +\infty.$$

Before proceeding with the discussion on the existence and characterization of the $\mathcal{D}_\phi$-divergence, we recall some basic concepts from convex analysis. For a detailed discussion of the concepts introduced below see Rockafellar (1970) and Hiriart-Urruty and Lemarechal (1993). The Legendre conjugate of the pair (C, ϕ) is the pair (C^*, ϕ^*), where C^* is the image of C under under the gradient mapping $\phi'(\cdot)$, and ϕ^* is the function on C^* given by:

$$\phi^*(\upsilon) \equiv \upsilon(\phi')^{-1}(\upsilon) - \phi\left((\phi')^{-1}(\upsilon)\right).$$

The following lemma establishes several useful properties of the Legendre conjugate ϕ^* of ϕ:

Lemma 1. *Under Assumption A6, we have: (i) $\phi^* \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$, (ii) ϕ^* is strictly convex, (iii) $\phi^* > 0$ on $\mathbb{R}^*_+$, (iv) ${\phi^*}' > 0$ on $\mathbb{R}$, (v) ${\phi^*}'(\upsilon) = (\phi')^{-1}(\upsilon)$ for any $\upsilon \in \mathbb{R}$, (vi) ${\phi^*}''(\upsilon) = [\phi''((\phi')^{-1}(\upsilon))]^{-1}$ for any $\upsilon \in \mathbb{R}$.*

We are now ready to discuss existence and uniqueness of the $\mathcal{D}_\phi$-projection of f_{t,s_i} characterized in Definition 2.

4.2. **Existence and Uniqueness.** Take the conditional moment constraint in Equation (13) which was used to define our semiparametric model. It is easy to show that the resulting set $\mathcal{SP}_{\theta,t,s_i}$ in (14) is *convex*. Even for nice convex sets such as $\mathcal{SP}_{\theta,t,s_i}$, however, the $\mathcal{D}_\phi$-projection of f_{t,s_i} onto $\mathcal{SP}_{\theta,t,s_i}$ may not exist.

In the literature (Teboulle and Vajda, 1993; Csiszar, 1995), the usual way to derive sufficient conditions for the existence of $g^*_{t,s_i}(\cdot,\theta)$ is to require that—in addition to being convex—the set $\mathcal{SP}_{\theta,t,s_i}$ be *closed*, say in L^1 norm. Such closedness condition would be satisfied if either the range of the structural mapping r in Equation (4) were bounded, which is equivalent to boundedness of the structural disturbances $U_{t,s}$, or if its domain were bounded, which is equivalent to boundedness of the endogenous variables $Y_{t,s}$. Both these conditions are contradictory to our setup, and our Assumption A1 explicitly rules out either possibility.

Our approach to establishing the existence of $g^*_{t,s_i}(\cdot,\theta)$ is based on the following intuitive argument: Under Assumption A3 we have that $f_{t,s_i} \in \mathcal{SP}_{\theta_0,t,s_i}$. In addition, $\mathcal{D}_\phi(f_{t,s_i}, f_{t,s_i}) = 0$ a.s.. Hence, at the true value θ_0 of the structural parameter θ we have, for any $T \geqslant 1$, any

$N \geqslant 1$, every $(t,i) \in [[1,T]] \times [[1,N]]$, and every $y \in \mathbb{R}^G$:

$$g^*_{t,s_i}(y,\theta_0) = f_{t,s_i}(y) \text{ a.s..} \tag{16}$$

In other words, when $\theta = \theta_0$ the $\mathcal{D}_\phi$-projection of f_{t,s_i} onto $\mathcal{SP}_{\theta_0,t,s_i}$ exists and is a.s. P unique. Provided we can invoke Implicit Function Theorem, it should then hold that for small deviations of θ around θ_0, the projection of f_{t,s_i} onto $\mathcal{SP}_{\theta,t,s_i}$ exists as well. We now provide a more formal treatment of this argument.

We start by restricting our attention to dynamic systems in Equation (4) in which r is continuously differentiable with respect to the structural parameter θ.

Assumption A7. For any $T \geqslant 1$, $N \geqslant 1$, and every $(\tilde{Y}'_{TN}, \tilde{X}'_{TN})' \in \mathbb{R}^{(G+K)TN}$, the mapping $\theta \mapsto r(Y_{t,s}, \dots, Y_{t-\tau,s}, X_{t,s}, \dots, X_{t-\tau,s}, \theta)$ is in $\mathcal{C}^1(\Theta, \mathbb{R}^G)$.

For structural mappings that satisfy Assumption A7 we let $D_\theta r_{t,s}(Y_{t,s}, \cdot) \in L(\mathbb{R}^k, \mathbb{R}^G)$ denote the partial derivative of $r_{t,s}$, defined previously, with respect to θ.

We first restrict the behavior of the Legendre conjugate ϕ^* and its derivative $\phi^{*'}$ by imposing several local integrability conditions. In what follows, $\mathcal{U}(\theta_0, \varepsilon) \equiv \mathcal{B}((\theta'_0, 0, 0')', \varepsilon)$ is an open ball in $\mathbb{R}^{k+G+1}$, centered at $(\theta'_0, 0, 0')'$ and with radius $\varepsilon > 0$.

Assumption A8. For any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, there exists $\mathcal{U}(\theta_0, \varepsilon_1) \subset \Theta \times \mathbb{R}^{G+1}$ such that for every $(\theta', \eta, \lambda')' \in \mathcal{U}(\theta_0, \varepsilon_1)$ we have: (i) $E\big[\phi^*\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s.; (ii) $E\big[\phi^{*'}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s.; (iii) $E\big[|r_{t,s_i}(Y_{t,s_i}, \theta)|\phi^{*'}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s..

Assumption A8 effectively imposes restrictions on the true conditional densities f_{t,s_i}. We now give an interpretation of A8(i,ii) in the case of I-divergence: $\phi(u) = u \ln u - u + 1$. The Legendre conjugate of ϕ then equals $\phi^*(v) = \exp v - 1$, so the properties in A8(i,ii) hold under a conditional version of a *weak Cramér condition*: for every θ in a neighborhood of θ_0 and every λ close to $0 \in \mathbb{R}^G$, we have $\int_{\mathbb{R}^G} \exp\big(\lambda' r_{t,s_i}(y,\theta)\big) f_{t,s_i}(y) dy < \infty$ a.s.. The Cramér condition restricts the generating function for the conditional moments of $r_{t,s_i}(Y_{t,s_i}, \theta)$—when

θ is close to θ_0—to be finite on a neighborhood of zero, at which the restriction is obviously satisfied.

Originally employed by Cramér (1938), the condition was imposed in order to apply a bound on the error of the normal approximation in the Central Limit Theorem (CLT) for univariate iid random variables. First multivariate extension of the Cramér condition can be found in Borovkov and Rogozin (1965). Alternatively, this condition can be interpreted as requiring that the conditional distribution of the disturbances in Equation (4) be smooth.

The following conditions ensure that one can differentiate under integral sign:

Assumption A9. For any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, there exists $\mathcal{U}(\theta_0, \varepsilon_2) \subset \Theta \times \mathbb{R}^{G+1}$ such that we have:
(i) $E\big[\sup_{(\theta',\eta,\lambda')' \in \mathcal{U}(\theta_0,\varepsilon_2)} \phi^{*''}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i},\theta)\big)\big(1 + |r_{t,s_i}(Y_{t,s_i},\theta)|^2\big)\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s.
(ii)$E\big[\sup_{(\theta',\eta,\lambda')' \in \mathcal{U}(\theta_0,\varepsilon_2)} \phi^{*''}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i},\theta)\big)\|D_\theta r_{t,s_i}(Y_{t,s_i},\theta)'\lambda r_{t,s_i}(Y_{t,s_i},\theta)'\|\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s.
(iii) $E\big[\sup_{(\theta',\eta,\lambda')' \in \mathcal{U}(\theta_0,\varepsilon_2)} \phi^{*''}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i},\theta)\big)|D_\theta r_{t,s_i}(Y_{t,s_i},\theta)'\lambda|\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s., and
(iv) $E\big[\sup_{(\theta',\eta,\lambda')' \in \mathcal{U}(\theta_0,\varepsilon_2)} \phi^{*'}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i},\theta)\big)\|D_\theta r_{t,s_i}(Y_{t,s_i},\theta)\|\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s..

Assumption A9 is used to ensure that Lebesgue Dominated Convergence Theorem applies, i.e. that we can interchange the order of integration and differentiation in the first order conditions that characterize the projection $g^*_{t,s_i}(\cdot,\theta)$ in Definition 2. In order to apply Implicit Function Theorem to those conditions obtained when $\theta = \theta_0$, we require the following invertibility assumption:

Assumption A10. For any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, we have $E\big[r_{t,s_i}(Y_{t,s_i},\theta_0) r_{t,s_i}(Y_{t,s_i},\theta_0)'\big|\mathcal{F}_{t,s_i}\big]$ invertible a.s..

We are now ready to state the main result of this section. As previously, $\mathcal{B}(\theta_0,\varepsilon)$ is an open ball in $\mathbb{R}^k$, centered at θ_0 and with radius $\varepsilon > 0$.

Theorem 1. *Let Assumptions A1-A4 and A6-A10 hold. Then, for any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, there exists $\mathcal{B}(\theta_0,\varepsilon) \subset \Theta$ such that for every $\theta \in \mathcal{B}(\theta_0,\varepsilon)$, the $\mathcal{D}_\phi$-projection $g^*_{t,s_i}(\cdot,\theta)$ of f_{t,s_i} onto $\mathcal{SP}_{\theta,t,s_i}$ exists, is P a.s. unique, and positive for a.e. ω.*

The projection $g^*_{t,s_i}(\cdot,\theta)$ *is given by:*

$$g^*_{t,s_i}(y,\theta) \equiv \phi^{*'}(\eta_{t,s_i}(\theta) + \lambda_{t,s_i}(\theta)' r_{t,s_i}(y,\theta)) f_{t,s_i}(y), \text{ for every } y \in \mathbb{R}^G, \tag{17}$$

where $(\eta_{t,s_i}(\theta), \lambda_{t,s_i}(\theta)) \equiv \arg\inf_{(\eta,\lambda')' \in \mathbb{R}^{G+1}} E\big[\phi^*\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i},\theta)\big)\big|\mathcal{F}_{t,s_i}\big] - \eta.$

We first comment on the strength of the assumptions used in Theorem 1. Csiszar (1995) gives a proof of the existence of the $\mathcal{D}_\phi$-projection that does not make topological assumptions on the set $\mathcal{SP}_{\theta,t,s_i}$ and/or on the range of the random variables $U_{t,s}$ and $Y_{t,s}$. In particular, Corollary to Theorem 3 in Csiszar (1995) is based on a moment condition on the convex conjugate ϕ^* of ϕ. Under I-divergence, this condition translates into a *strong Cramér condition*, whereby "strong" we mean that the finiteness of the generating function for the conditional moments of $r_{t,s_i}(Y_{t,s_i},\theta)$ (when θ is close to θ_0) needs to hold for all $\lambda \in \mathbb{R}^G$. This condition is obviously stronger than our "weak" version imposed in Assumption A8, which only needs to hold for λ in some neighborhood of $0 \in \mathbb{R}^G$.

Theorem 1 establishes two important results. First, it shows that the $\mathcal{D}_\phi$-projection $g^*_{t,s_i}(\cdot,\theta)$ of f_{t,s_i} onto $\mathcal{SP}_{\theta,t,s_i}$ exists and is unique, that is except for the unavoidable nonuniqueness within a P-equivalence class of random variables. As pointed out previously, this result exploits the existence of the $\mathcal{D}_\phi$-projection when $\theta = \theta_0$ and extends it by means of Implicit Function Theorem. It is worth noting that the proof of Theorem 1—in particular, the part of the proof corresponding to Lemma 2—establishes in a direct way that there exists $g^*_{t,s_i}(\cdot,\theta)$ in $\mathcal{SP}_{\theta,t,s_i}$ with density (17). An early suggestion of such direct approach can be found in Csiszar (1975) (see a discussion on p.156 in Csiszar (1975)).

The second key result of Theorem 1 is to derive the analytic expression of $g^*_{t,s_i}(\cdot,\theta)$. The density of the $\mathcal{D}_\phi$-projection obtained in Equation (17) reveals an interesting property: it is parameterized by two random *finite dimensional* Lagrange multipliers $\eta_{t,s_i}(\theta)$ and $\lambda_{t,s_i}(\theta)$, both of which are $\mathcal{F}_{t,s_i}$-measurable and depend on θ. In other words, projecting onto the semiparametric set $\mathcal{SP}_{\theta,t,s_i}$ reduces the problem to the one in which, for any $y \in \mathbb{R}^G$, the density $g^*_{t,s_i}(y,\theta)$ can be written as a product of two terms: a first one $\phi^{*'}(\eta_{t,s_i}(\theta) + \lambda_{t,s_i}(\theta)' r_{t,s_i}(y,\theta))$

that is finitely parameterized by θ, and a second one that is the true density $f_{t,s_i}(y)$ which does not depend on θ. We now derive additional useful properties of $g^*_{t,s_i}(\cdot,\theta)$.

4.3. **Definition and Properties of $\mathcal{P}^*$.** A number of interesting properties can be derived from the expression of the $\mathcal{D}_\phi$-projected density $g^*_{t,s_i}(\cdot,\theta)$ obtained in Theorem 1. We start by restricting our attention to local parameter sets Θ' defined as follows.

Definition 3. Let $\mathcal{B}(\theta_0,\bar{\varepsilon})$ be the largest open ball under which the results of Theorem 1 hold for every $T \geqslant 1$ and $N \geqslant 1$. Then Θ' is the largest compact set contained in $\mathcal{B}(\theta_0,\bar{\varepsilon})$.

In general $\Theta' \subset \Theta$ unless Theorem 1 holds for $\mathcal{B}(\theta_0,\bar{\varepsilon}) = \mathrm{Int}(\Theta)$. Notice that restricting the parameter set to Θ' is without loss of generality. Since our goal is to derive a semiparametric bound for θ_0, we can limit our analysis to a neighborhood of θ_0.

We now use Θ' to define our parametric submodel $\mathcal{P}^*$. By construction, for any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, we have $g^*_{t,s_i}(\cdot,\theta) \in \mathcal{SP}_{\theta,t,s_i}$; combining this with the property in Equation (16) then shows, by Definition 1, the following Corollary to Theorem 1.

Corollary 2. *Assume the conditions of Theorem 1 hold, and let Θ' be as in Definition 3. Then, for any $T \geqslant 1$ and any $N \geqslant 1$, $\mathcal{P}^* \equiv \bigcup_{\theta\in\Theta'} \mathcal{P}^*_\theta$ with $\mathcal{P}^*_\theta \equiv \bigotimes_{t\in[[1,T]]} \bigotimes_{i\in[[1,N]]} \mathcal{P}_{\theta,t,s_i}$ and $\mathcal{P}^*_{\theta,t,s_i} \equiv \left\{g^*_{t,s_i}(\cdot,\theta)\right\}$ is a parametric submodel of $\mathcal{SP}$.*

In other words, Corollary 2 says that the model $\mathcal{P}^*$ satisfies all the conditions of Definition 1. The next step needed to derive the infimum (i.e. the greatest lower bound) of the semiparametric efficiency bound for θ_0 is to derive the asymptotic variance of the MLE in the parametric submodel $\mathcal{P}^*$. We shall see that the latter depends on the derivatives with respect to θ of the $\mathcal{D}_\phi$-projections $g^*_{t,s_i}(\cdot,\theta)$, which we study next.

Similar to the identity derived in Equation (16), we are now interested in the values that successive derivatives of $g^*_{t,s_i}(\cdot,\theta)$ with respect to the structural parameter θ (when they exist) take at the true value θ_0. Under the same set of conditions as in Theorem 1, we have the following result:

Corollary 3. *Assume the conditions of Theorem 1 hold and that Θ' is as in Definition 3. Then, for any $T \geqslant 1$, $N \geqslant 1$, and every $(t,i) \in [[1,T]] \times [[1,N]]$, the Lagrange multipliers $\eta_{t,s_i}(\theta)$ and $\lambda_{t,s_i}(\theta)$ are continuously differentiable on Θ' a.s. with: $D_\theta \eta_{t,s_i}(\theta_0) = 0$ a.s., and $D_\theta \lambda_{t,s_i}(\theta_0) = E\big[D_\theta r_{t,s_i}(Y_{t,s_i},\theta_0)\big|\mathcal{F}_{t,s_i}\big]\big\{E\big[r_{t,s_i}(Y_{t,s_i},\theta_0) r_{t,s_i}(Y_{t,s_i},\theta_0)'|\mathcal{F}_{t,s_i}\big]\big\}^{-1}$ a.s..*

In particular, Corollary 3 implies that the least favorable densities $g^*_{t,s_i}(\cdot,\theta)$ in Theorem 1 are continuously differentiable with respect to θ with scores that satisfy:

$$D_\theta \ln g^*_{t,s_i}(y,\theta_0) =$$

$$r_{t,s_i}(y,\theta_0)\big\{E\big[r_{t,s_i}(Y_{t,s_i},\theta_0) r_{t,s_i}(Y_{t,s_i},\theta_0)'|\mathcal{F}_{t,s_i}\big]\big\}^{-1} E\big[D_\theta r_{t,s_i}(Y_{t,s_i},\theta_0)|\mathcal{F}_{t,s_i}\big], \tag{18}$$

for every $y \in \mathbb{R}^G$. The assumptions of Theorem 1 (and its Corollaries 2 and 3) will be required for the asymptotic analysis of the MLE based on $\mathcal{P}^*$, to which we turn next.

5. Least Favorable Maximum Likelihood Estimator

We now derive the asymptotic distribution of the maximum likelihood estimator (MLE) of θ_0 based on the least favorable densities $g^*_{t,s_i}(\cdot,\theta)$. The estimator is not feasible because the parameters $\lambda_{t,s_i}(\theta)$ and $\eta_{t,s_i}(\theta)$ of the least favorable distributions $g^*_{t,s_i}(\cdot,\theta)$ are generally unknown. The estimation procedure is a device to obtain the semiparametric bound under Assumption A3.

To avoid technical difficulties, we consider the case for which the structural mapping r in the dynamic system (4) is twice continuously differentiable with respect to all of its arguments.

Assumption A11. The mapping r is in $\mathcal{C}^2(\mathbb{R}^{(G+K)\tau} \times \Theta', \mathbb{R}^G)$.

5.1. **Estimator.** We start by defining the least favorable MLE of θ_0. For this, start by fixing $T \geqslant 1$ and $N \geqslant 1$, and consider the product of conditional densities $g^*_{t,s_i}(\cdot,\theta)$ obtained as:

$$g^*_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{y}_{TN},\tilde{x}_{TN},\theta) \equiv \prod_{t=1}^{T}\prod_{i=1}^{N} g^*_{t,s_i}(y_{t,s_i},\theta) \tag{19}$$

where the labeling of the elements in $\mathbb{S}_N$ is done as before, i.e. $s_1 \in \mathbb{S}_1$, $s_2 \in \mathbb{S}_2 \backslash \mathbb{S}_1$, up to $s_N \in \mathbb{S}_N \backslash \mathbb{S}_{N-1}$. Under assumptions of Theorem 1, by Corollary 2, we know that the expression in Equation (19) defines a model for the conditional density for $\tilde{Y}_{TN}$ given $\tilde{X}_{TN}$ that is parameterized by θ, and correctly specified, i.e. $g^*_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{Y}_{TN}, \tilde{X}_{TN}, \theta_0) = f_{\tilde{Y}_{TN}|\tilde{\tilde{X}}_{TN}}(\tilde{Y}_{TN})$ a.s.. Moreover, under Assumption A5, for any $(\theta_1, \theta_2) \in \Theta'$ we have: $g^*_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{Y}_{TN}, \tilde{X}_{TN}, \theta_1) = g^*_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{Y}_{TN}, \tilde{X}_{TN}, \theta_2)$ a.s. only if $\theta_1 = \theta_2$, so the model in (19) is in addition identified on Θ'.

Let then $\mathcal{L}(\theta|\tilde{x}_{TN}, \tilde{y}_{TN})$ denote the likelihood function:

$$\mathcal{L}(\theta|\tilde{x}_{TN}, \tilde{y}_{TN}) = g^*_{\tilde{Y}_{TN}|\tilde{\mathcal{X}}_{TN}}(\tilde{y}_{TN}, \tilde{x}_{TN}, \theta) f_{\tilde{X}_{TN}}(\tilde{x}_{TN}) \tag{20}$$

where $f_{\tilde{X}_{TN}}$ denotes the joint distribution of the exogenous variables $\tilde{X}_{TN}$. From the separability assumption A2 we know that the latter does not depend on θ. Using the Kullback-Leibler inequality it follows that for any $\theta \in \Theta'$, we have $E\big[\ln \mathcal{L}(\theta_0|\tilde{X}_{TN}, \tilde{Y}_{TN})\big] \geqslant E\big[\ln \mathcal{L}(\theta|\tilde{X}_{TN}, \tilde{Y}_{TN})\big]$, with the expectation being taken with respect to the joint measure of $\tilde{Y}_{TN}$ and $\tilde{X}_{TN}$. Combining the above inequality with Equations (20) and (19) then shows that θ_0 solves the optimization problem:

$$\max_{\theta \in \Theta'} E\left[\sum_{t=1}^{T} \sum_{i=1}^{N} \ln g^*_{t,s_i}(Y_{t,s_i}, \theta)\right].$$

From the expression of the least favorable densities obtained in Theorem 1, solving the above optimization problem is equivalent to solving:

$$\max_{\theta \in \Theta'} E\left[\sum_{t=1}^{T} \sum_{i=1}^{N} G_{t,s_i}(\theta)\right], \text{ with } G_{t,s_i}(\theta) \equiv \ln \phi^{*'}(\eta_{t,s_i}(\theta) + \lambda_{t,s_i}(\theta)' r_{t,s_i}(Y_{t,s_i}, \theta)) \tag{21}$$

which suggests estimating θ_0 by solving the sample counterpart:

$$\max_{\theta \in \Theta'} \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} G_{t,s_i}(\theta). \tag{22}$$

5.2. **Asymptotic Distribution.** We consider the asymptotic properties of the least favorable MLE θ_{TN} obtained by solving the maximization problem in Equation (22) when both T and N to tend to infinity; see Hahn and Kuersteiner (2002, 2004), for example.

There are three schemes under which the asymptotic distribution of estimators can be derived when $(T, N) \to \infty$. The first is by using the sequential limit approach composed of two steps: in the first step, N is fixed and T is allowed to pass to infinity; in the second, N is let to pass to infinity. Another approach is to pass to infinity along a specific diagonal path determined by a function relation of the type $T = T(N)$ with $N \to \infty$. A third approach is to let N and T to pass to infinity simultaneously. The joint limit requires stronger conditions, but it does not run into the drawback of the other two approaches. Phillips and Moon (1999), for example, give an exhaustive discussion on the drawbacks of sequential and diagonal limits.

To deal with the joint limit theory without restricting the cross-sectional dependence we use the concept of *random fields*.[16] The collection $Z = \{Z_{t,s}, (t,s) \in \mathbb{N} \times \mathbb{S}\}$ of random vectors in $\mathbb{R}^{K+G}$ composed of the explanatory and dependent variables:

$$Z_{t,s} \equiv (X'_{t,s}, Y'_{t,s})' \tag{23}$$

for any $(t,s) \in \mathbb{N} \times \mathbb{S}$, defines a vector random field. As previously, $(\mathbb{R}^{K+G})^{\mathbb{N}\times\mathbb{S}}$ denotes the countable product space, and $Z : \Omega \to (\mathbb{R}^{K+G})^{\mathbb{N}\times\mathbb{S}}$. For any subset of indices $V \subseteq \mathbb{N} \times \mathbb{S}$, we introduce the σ-algebra $\mathcal{Z}_V \equiv \sigma(Z_{t,s}, (t,s) \in V)$.

We limit our attention to random fields Z that are α-mixing (or "strong" mixing). Mixing conditions for random field extend the usual mixing conditions on random sequences. Rosenblatt (1986) gives the following definition.

Definition 4 (**Strong Mixing**)**.** Let $V \subseteq \mathbb{N} \times \mathbb{S}$ and $W \subseteq \mathbb{N} \times \mathbb{S}$ be two sets of indices, and let $d(V, W)$ be a distance between them. Consider $Z_V \equiv (Z_v, v \in V)$ and $Z_W \equiv (Z_w, w \in W)$, and the associated σ-algebras $\mathcal{Z}_V$ and $\mathcal{Z}_W$. The random field Z is said to be α-mixing (or strong mixing) if there exists a function φ satisfying $\varphi(d) \to 0$ as $d \to \infty$, for which:

$$\alpha(\mathcal{Z}_V, \mathcal{Z}_W) \equiv \sup_{A \in \mathcal{Z}_V,\, B \in \mathcal{Z}_W} [P(Z_V \in A \cap Z_W \in B) - P(Z_V \in A)P(Z_W \in B)] \leqslant \varphi(d(V, W)).$$

[16] Ivanov and Leonenko (1986) and Föllmer (1988) provide a detailed analysis of random fields.

For nonempty sets $V \subseteq \mathbb{N} \times \mathbb{S}$ and $W \subseteq \mathbb{N} \times \mathbb{S}$ that are *disjoint*, we use the abbreviation $\alpha(V, W) \equiv \alpha(\mathcal{Z}_V, \mathcal{Z}_W)$. For any $m \in \mathbb{N}$ and $(k, l) \in \mathbb{N} \cup \{\infty\}$, the mixing coefficients for the random field Z are defined as:

$$\alpha_{k,l}(m) \equiv \sup\{\alpha(V, W) : |V| \leqslant k, |W| \leqslant l, d(V, W) \geqslant m\} \tag{24}$$

where $|V|$ and $|W|$ denote the cardinalities of the sets V and W, respectively. The mixing coefficient $\alpha_{k,l}(m)$ was introduced by Bulinsky and Zhurbenko (1976) (see also Bolthausen, 1982). Notice that the dependence of the σ-algebras $\mathcal{Z}_V$ and $\mathcal{Z}_W$ may increase as the sets V and W become larger if the distance between them is not preserved. The role of $\alpha_{k,l}(m)$ is to quantify how this dependence diminishes as the sets V and W become more distant and their cardinalities do not exceed given values k and l.

We are now ready to state the *dependence* condition imposed on the random field Z. Hereafter, we shall make the following assumption:

Assumption A12. The vector random random field $Z = \{Z_{t,s}, (t, s) \in \mathbb{N} \times \mathbb{S}\}$, with $Z_{t,s} = (X'_{t,s}, Y_{t,s})' \in \mathbb{R}^{K+G}$ is α-mixing with mixing coefficients satisfying:

(a) $\sum_{m=1}^{\infty} m\, \alpha_{k,l}(m) < \infty$, for $k + l \leqslant 4$;

(b) $\sum_{m=1}^{\infty} m\, \alpha_{1,1}(m)^{\delta/(2+\delta)} < \infty$, for some $\delta > 0$;

(c) $\alpha_{1,\infty}(m) = O(m^{-2-\varepsilon})$.

Assumption A12 it is generally not sufficient to guarantee that nonlinear functions of the random field Z, such as the objective function G_{t,s_i} in Equation (22), satisfy a law of large numbers and a central limit theorem. The difficulty stems from the fact that the Lagrange multipliers $\eta_{t,s_i}(\theta)$ and $\lambda_{t,s_i}(\theta)$ are functions of an increasing number of vectors in the random field Z as T and N get large. In order to avoid such difficulties, we limit the dependence of $\eta_{t,s_i}(\theta)$ and $\lambda_{t,s_i}(\theta)$ on some finite number of vectors in Z. We first introduce the following set of indices: $\Pi_{t,s_i} \equiv \{(h, s_j) : |h - t| \leqslant \kappa, |j - i| \leqslant \varsigma\}$, where $\kappa \in \mathbb{N}$ and $\varsigma \in \mathbb{N}$ are maximal numbers of lags (and leads) and neighbors, respectively, which we assume to be finite: $\kappa < \infty$ and $\varsigma < \infty$. Let then $\tilde{Z}_{t,s_i} \equiv \{Z_{h,v}, (h, v) \in \Pi_{t,s_i}\}$ and $\mathcal{F}^{\Pi}_{t,s_i} \equiv \sigma(Y_{t,v}, v \in$

$\{s_{i-\varsigma},\dots,s_{i-1}\}, Y_{h,v},(h,v)\in[[t-\tau,t-1]]\times\{s_{i-\varsigma},\dots,s_{i+\varsigma}\}, X_{h,v},(h,v)\in\Pi_{t,s_i})$. Put in words, $\mathcal{F}^{\Pi}_{t,s_i}$ is the σ-algebra of those generating variables in $\mathcal{F}_{t,s_i}$ whose indices are restricted to lie in the set Π_{t,s_i}.

We make the following assumption.

Assumption A13. For every $(t,i,\theta)\in\mathbb{N}^2\times\Theta'$, $\eta_{t,s_i}(\theta)$ and $\lambda_{t,s_i}(\theta)$ are $\mathcal{F}^{\Pi}_{t,s_i}$-measurable.

Assumption A13 limits the dependence of the parameters of the least favorable distribution $g^*_{t,s_i}(\cdot,\theta)$ on a finite number of vectors in the random field Z.[17] This assumption is for instance satisfied if for every $\theta\in\Theta'$ and every $(\eta,\lambda')'$ in a neighborhood of $0\in\mathbb{R}^{G+1}$, we have: $E\big[\phi^*(\eta+\lambda' r_{t,s_i}(Y_{t,s_i},\theta))|\mathcal{F}_{t,s_i}\big]=E\big[\phi^*(\eta+\lambda' r_{t,s_i}(Y_{t,s_i},\theta))|\mathcal{F}^{\Pi}_{t,s_i}\big]$ a.s. (see Theorem 1). This property implies that the sequence of Lagrange multipliers—and hence the sequence of objective functions $\{G_{t,s_i},(t,i)\in\mathbb{N}^2\}$—are α-mixing if the underlying random field Z is α-mixing. Replacing Assumption A13 with any assumption that guarantees α-mixing of $\{\eta_{t,s_i}(\theta),(t,i)\in\mathbb{N}^2\}$ and $\{\lambda_{t,s_i}(\theta),(t,i)\in\mathbb{N}^2\}$ would not affect the results of this section. As such, Assumption A13 is to be regarded as purely technical requirement that does not impact the generality of our results.

The derivation of consistency for the estimator defined in Equation (22) involves a demonstration that the objective function of the estimator converges uniformly over Θ' to its asymptotic counterpart. This is tantamount to showing that the uniform law of large numbers (ULLN) holds for $\{G_{t,s_i}(\theta),(t,i)\in\mathbb{N}^2\}$, i.e. that as $(T,N)\to\infty$ we have, uniformly on $[[1,T]]\times[[1,N]]$,

$$\sup_{\theta\in\Theta'}\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}G_{t,s_i}(\theta)-E\big[G_{t,s_i}(\theta)\big]\right|\xrightarrow{p}0.$$

ULLN can be established is by imposing conditions which transform a pointwise convergence result delivered by an appropriate LLN into the corresponding uniform one. Here, we follow recent work by Jenish and Prucha (2007) to give sufficient conditions for ULLN for possibly nonstationary random fields. Two key assumptions are sufficient for a ULLN to hold

[17]Note that under Assumption A13, the Lagrange multipliers no longer depend on T and N. In other words, they no longer form a triangular array of random fields. For this reason, the qualifier "for any $T\geqslant 1$, $N\geqslant 1$, and every $(t,i)\in[[1,T]]\times[[1,N]]$" can be reduced to "for every $(t,i)\in\mathbb{N}^2$".

in our context. First, a (pointwise) LLN holds for all $\theta \in \Theta'$. Second, $\{G_{t,s_i}(\theta), (t,i) \in \mathbb{N}^2\}$ is L_0 stochastically equicontinuous on Θ', i.e. for every $\epsilon > 0$

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} P\left(\sup_{\theta' \in \Theta'}\sup_{\theta \in \mathcal{B}(\theta',\delta)}\left|G_{t,s_i}(\theta) - G_{t,s_i}(\theta')\right| > \epsilon\right) \to 0, \quad \text{as } \delta \to 0.$$

The following assumptions are sufficient to establish consistency of the MLE θ_{TN}.

Assumption A14. (i) $\sup_{(t,i)\in\mathbb{N}^2} E\left[|\sup_{\theta\in\Theta'} G_{t,s_i}(\theta)|^{1+\delta}\right] < \infty$, for δ in Assumption A12(b); (ii) for every $(t,i) \in \mathbb{N}^2$ and every $(\theta, \theta') \in (\Theta')^2$, we have: $|\,G_{t,s_i}(\theta) - G_{t,s_i}(\theta')| \leqslant B_{t,s_i}|\,\theta - \theta'|$ a.s. where $\{B_{t,s_i}, (t,i) \in \mathbb{N}^2\}$ satisfies $\lim_{(T,N)\to\infty}(TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} E(B_{t,s_i}^p) < \infty$ for some $p > 0$.

Assumption A14(i) bounds the moments of the objective function. The Lipschitz property in A14(ii) implies that $\{G_{t,s_i}, (s,i) \in \mathbb{N}^2\}$ is L_0 equicontinuous on Θ'.

Theorem 4. *Let Assumptions A1-A11,A13-A14 be satisfied. Moreover, let Assumption A12(a) hold for $k = l = 1$. Then a solution θ_{TN} to the maximization problem in Equation* (22) *exists, and $\theta_{TN} \xrightarrow{p} \theta_0$ as $(T,N) \to \infty$.*

The proof of Theorem 4 uses a uniform law of large numbers for random fields, established by Jenish and Prucha (2007). Note that there are no requirements on the relative speeds with which the two sample sizes increase, i.e. we can have $T/N \to c$ with $c \in \mathbb{R} \cup \{\infty\}$.

Recalling that $g^*_{t,s_i}(\cdot,\theta)$ is twice continuously differentiable with respect to θ, let $H_\theta \ln g^*_{t,s_i}(\cdot,\theta)$ denote the Hessian of the log-likelihood and define the expected Hessian evaluated at θ_0 as: $H \equiv E\left[H_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta_0)\right]$. To obtain asymptotic normality, we define:

$$V \equiv \lim_{(T,N)\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N} D_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta_0)\right)$$

and seek conditions which will ensure the unit asymptotic normality of:

$$(25)\qquad \frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N} \upsilon' V^{-1/2} D_\theta \ln g^*_{t,s_i}(y_{t,s_i},\theta_0)$$

for arbitrary vectors $\upsilon \in \mathbb{R}^K$ such that $\upsilon'\upsilon = 1$. Using the Cramér-Wold device, this will establish that that $(TN)^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N} V^{-1/2} D_{\theta_0} g^*_{t,s_i}(Y_{t,s_i},\theta_0) \xrightarrow{d} N(0, I_k)$. In our setting, asymptotic normality of the quantity defined in Equation (25) can be deduced from Jenish and Prucha's (2007) CLT for random fields. Uniform convergence and non-singularity of the Hessian will establish the asymptotic distribution of the MLE.

Before stating the assumptions under which the asymptotic normality for θ_{TN} is proved, it is helpful to examine more closely the the form of the gradient $D_\theta \ln g^*_{t,s_i}(\cdot,\theta)$ evaluated at $\theta = \theta_0$. Under the assumptions of Theorem 1, using the law of iterated expectations and the expression derived in Equation (18), it follows that for every $(t,i) \in \mathbb{N}^2$, we have: $E\big[D_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta_0)\big] = 0$ a.s..

The following primitive assumptions are sufficient for asymptotic normality of θ_{TN}.

Assumption A15. For every $(t,i) \in \mathbb{N}^2$, we have: (i) $E\big[|\, D_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta_0)|^{2+\delta}\big] < \infty$, with δ as in Assumption A12(b); (ii) for every $\theta \in \Theta'$, $E\big[|\, H_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta)|^{1+\xi}\big] < \infty$ for some $\xi > 0$ and $|H_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta') - H_\theta \ln g^*_{t,s_i}(Y_{t,s_i},\theta)| \leqslant C_{t,s_i}(\theta)|\,\theta' - \theta|$ a.s. where $\{C_{t,s_i}, (t,i) \in \mathbb{N}^2\}$ satisfies $\lim_{(T,N)\to\infty}(TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} E\big[C_{t,s_i}\big] < \infty$.

Assumption A15(i) guarantees that the appropriate version of the CLT can be applied to the quantity in Equation (25) and that the Hessian converge uniformly to a positive definite matrix. Underlying the existence of the Hessian of the objective functions are regularity conditions on certain moments of $r_{t,s_i}(Y_{t,s_i},\theta)$ and its derivatives that guarantee that $\eta_{t,s_i}(\theta)$ and $\lambda_{t,s_i}(\theta)$ are twice differentiable on Θ'. Assumption A15(ii) imposes additional smoothness on the Hessian.

The expected value of the Hessian at θ_0 can be calculated explicitly by twice differentiating the objective function $G_{t,s_i}(\theta)$ in Equation (21) with respect to θ. However, correct specification of the parametric model $\mathcal{P}^*$ allows an alternative derivation of the expected Hessian at θ_0. By the information matrix equality we have:

$$\begin{aligned}(26)\quad H = \lim_{(T,N)\to\infty} \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} E\Big[& E\big[D_\theta r_{t,s_i}(Y_{t,s_i},\theta_0))'|\mathcal{F}^{\Pi}_{t,s_i}\big] \\ & \times E\big[r_{t,s_i}(Y_{t,s_i},\theta_0) r_{t,s_i}(Y_{t,s_i},\theta_0)'|\mathcal{F}^{\Pi}_{t,s_i}\big]^{-1} E\big[D_\theta r_{t,s_i}(Y_{t,s_i},\theta_0))|\mathcal{F}^{\Pi}_{t,s_i}\big]\Big].\end{aligned}$$

Note that—in virtue of Assumption A13—the conditioning σ-algebra in the expression of the Hessian is $\mathcal{F}_{t,s_i}^{\Pi}$ and not $\mathcal{F}_{t,s_i}$.

We have the following asymptotic normality result:

Theorem 5. *Let Assumptions A1-A11, A13-A15 hold. Further let Assumption A12 be satisfied for some $\delta, \varepsilon > 0$. Assume that for any $T \geqslant 1$ and $N \geqslant 1$, we have $\theta_{TN} \in \text{int}(\Theta')$. Then, $\sqrt{TN}(\theta_{TN} - \theta_0) \xrightarrow{d} N(0, H^{-1})$.*

The MLE estimator based on the projection of $f_{Y_{t,i}|\mathcal{F}_{t,s_i}}$ on $\mathcal{SP}_{\theta,t,s_i}$ is consistent and asymptotically normal. The projected densities form a parametric submodel and hence no semiparametric estimator can have asymptotic variance smaller (in the semidefinite sense) than H^{-1}.

6. Conclusion

We have derived an efficiency bound for structural cross section time series models that come in a form of a system of dynamic nonlinear equations. The models we consider exhibit both temporal and spatial dependence, and heterogeneity among variables. We construct a (least favorable) parametric submodel of a semiparametric model defined by a panel exogeneity condition. This condition generalizes the notion of strict stationarity used in setups with independent and identically distributed variables. The variance of asymptotic distribution of the estimator of the parameter of interest gives a bound on the efficiency bound: no semiparametric estimator can have variance lower than H^{-1}. The lower bound is derived under the large T large N thought experiment. Asymptotic results are established by using a law of large numbers and a central limit theorem for random fields recently given in Jenish and Prucha (2007). Our results extend to a setting with the T fixed, in which case the bound can be derived by assuming a richer time dependence structure.

Constructing an estimator possessing asymptotic variance equal to H^{-1} is the next step in the analysis of the efficiency problem. We briefly consider what is viable strategy for the construction of such estimator. Let $\bar{\theta}_{TN}$ a consistent estimator of θ_0. This preliminary estimator can be easily obtained by using the unconditional moment restrictions generated by (13). Define the following weighting

$$\omega_{j,\ell}^{t,s_i} = \frac{K(\psi_{j,\ell}^{t,s_i})}{\sum_{j,\ell} K(\psi_{j,\ell}^{t,s_i}))}, \quad \psi_{j,\ell}^{t,s_i} = \frac{q_{t,s_i} - q_{j,s_\ell}}{h_{NT}}, \quad q \in \mathcal{F}_{t,s_i}^{\Pi},$$

where K is a kernel function, and h_{TN} denotes a null sequence of positive number such that a $TNh_{TN} \to \infty$. Consider the following minimization problem

$$\min_{\theta\in\Theta}\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} r_{t,s_i}^{w}(\theta)\right)' W_{T,N}\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} r_{t,s_i}^{w}(\theta)\right) \tag{27}$$

where

$$r_{t,s_i}^{w}(\theta) = \sum_{j=1}^{T}\sum_{\ell=1}^{N}\omega_{j,\ell}^{t,s_i} r_{t,s_i}(Y_{t,s_i},\theta),$$

$$W_{T,N} = \left[\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{T}\sum_{\ell=1}^{N}\omega_{j,\ell}^{t,s_i} r_{t,s_i}(Y_{t,s_i},\bar\theta) r_{t,s_i}(Y_{t,s_i},\bar\theta)'\right]^{-1}.$$

The role of the weighting is to incorporate the information about the conditional moment restrictions. Under suitable regularity conditions we have that $(TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} r_{t,s_i}^{\omega}(\theta) \xrightarrow{p} E\big[r_{t,s_i}(Y_{t,s_i},\theta)|\mathcal{F}_{t,s_i}^{\Pi}\big]$, $W_{T,N} \xrightarrow{p} W \equiv (TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} E\big[r_{t,s_i}(Y_{t,s_i},\theta_0) r_{t,s_i}(Y_{t,s_i},\theta_0)'|\mathcal{F}_{t,s_i}^{\Pi}\big]$, and, more generally, sample averages weighted by $\omega_{j,\ell}^{t,s_i}$ converge to their conditional (with respect to $\mathcal{F}_{t,s_i}^{\Pi}$) expected values. Mean value expansion of the first order conditions of the solution of (27), $\hat\theta$, gives

$$\begin{aligned} 0 = &\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} D_\theta r_{t,s_i}^{w}(\hat\theta)\right)' W_{T,N}\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} r_{t,s_i}^{w}(\theta_0)\right) \\ &+ \left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} D_\theta r_{t,s_i}^{w}(\hat\theta)\right)' W_{T,N}\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} D_\theta r_{t,s_i}^{w}(\dot\theta)\right)(\hat\theta - \theta_0). \end{aligned}$$

Under regularity conditions—such as uniform convergence of higher order moments of the weighted version of $r_{t,s_i}(Y_{t,s_i},\theta)$—the following approximation may be shown to hold

$$\sqrt{TN}\big(\hat\theta - \theta_0\big) = H^{-1}\left(\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N} E\big[D_\theta r_{t,s_i}(Y_{t,s_i},\theta_0)\big]\right) W^{-1}\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} r_{t,s_i}^{w}(\theta_0)\right) + o_p(1).$$

A version of the central limit theorem for random fields applied to the sample average of r^{ω}_{t,s_i} gives that

$$\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N} r^{w}_{t,s_i}(\theta_0) \xrightarrow{d} N(0,W).$$

By standard asymptotic arguments it follows that $\hat{\theta}$ is asymptotically normal with variance equal to the bound given in Theorem 5, H^{-1}.

Appendix A. Proofs

Proof of Proposition 1. Under A1(i)-(ii), the mapping $\tilde{r}(\cdot, \tilde{X}_N, \theta)$ is in $\mathcal{C}^1(\mathbb{R}^{GTN}, \mathbb{R}^{GTN})$, and its Jacobian $J(\cdot, \tilde{X}_N, \theta)$ never vanishes. Now, consider $y = (y_{t,s}, (t,s) \in [[1,T]] \times \mathbb{S}_N) \in \mathbb{R}^{GTN}$ ($y_{t,s} \in \mathbb{R}^G$ for any $(t,s) \in [[1,T]] \times \mathbb{S}_N$): then $|y|^2 = \sum_{t=1}^{T} \sum_{s \in \mathbb{S}_N} |y_{t,s}|^2$ and $|y| \to \infty$ implies that for least one (t_0, s_0), $(t_0, s_0) \in [[1,T]] \times \mathbb{S}_N$, we have $|y_{t_0,s_0}| \to \infty$. Assumption A1(iii) then ensures that at least for one t_1, $t_0 \leqslant t_1 \leqslant \max\{t_0 + \tau, T\}$, we have $|r(y_{t_1,s_0}, \ldots, y_{t_1-\tau,s_0}, X_{t_1,s_0}, \ldots, X_{t_1-\tau,s_0}, \theta)| \to \infty$. Given that $|\tilde{r}(y, \tilde{X}_N, \theta)|^2 = \sum_{t=1}^{T} \sum_{s \in \mathbb{S}_N} |r(y_{t,s}, \ldots, y_{t-\tau,s}, X_{t,s}, \ldots, X_{t-\tau,s}, \theta)|^2$, we then have that $|\tilde{r}(y, \tilde{X}_N, \theta)| \to \infty$; hence the mapping $\tilde{r}(\cdot, \tilde{X}_N, \theta)$ is proper. We can now apply Corollary 4.3 in Palais (1959), to show that $\tilde{r}(\cdot, \tilde{X}_N, \theta)$ is a diffeomorphism of $\mathbb{R}^{GTN}$ onto itself. Let then $q(\cdot, \tilde{X}_N, \theta)$ be its inverse, i.e. for any $(y, u) \in \mathbb{R}^{2GTN}$ we have: $q(u, \tilde{X}_N, \theta) = y \Leftrightarrow \tilde{r}(y, \tilde{X}_N, \theta) = u$, and $q(\cdot, \tilde{X}_N, \theta)$ is a diffeomorphism of $\mathbb{R}^{GTN}$ onto itself. □

Proof of Lemma 1. First, note that from the definition of the Legendre conjugate, ϕ^* is continuous and differentiable on $\mathbb{R}$. In addition, the derivative of ϕ^* is given by:

$$\phi^{*'}(\upsilon) = (\phi')^{-1}(\upsilon), \text{ for any } \upsilon \in \mathbb{R}.$$

Given the strict convexity of ϕ in Assumption A6(ii), ϕ' is continuous and strictly increasing on $\mathbb{R}_+^*$ with $\phi'(0) = -\infty$ from A6(v), and $\lim_{u \to +\infty} \phi'(u) = +\infty$ from A6(iv); so its inverse $\phi^{*'}$ is continuous and strictly increasing on $\mathbb{R}$. Hence, ϕ^* is strictly convex. Since $\lim_{\upsilon \to -\infty} \phi^{*'}(\upsilon) = 0$, we have $\phi^{*'} > 0$ in $\mathbb{R}$. Moreover, from A6(iii) $\phi^*(0) = 0$ which combined with the previous property gives $\phi^* > 0$ on $\mathbb{R}_+^*$. Finally, A6(ii) implies $\phi'' > 0$ on $\mathbb{R}_+^*$ so $\phi^{*'}$ is continuously differentiable on $\mathbb{R}$ with derivative:

$$\phi^{*''}(\upsilon) = \frac{1}{\phi''((\phi')^{-1}(\upsilon))}.$$

This completes the proof of Lemma 1. □

Proof of Theorem 1. The proof is done in two steps. In the first step, we show that the conclusion of Theorem 1 holds when $\theta = \theta_0$. In the second step, we invoke Implicit Function Theorem around the first order condition satisfied by the lagrange multipliers $(\eta_{t,s_i}(\theta_0), \lambda_{t,s_i}(\theta_0))$ to extend the results to a neighborhood of θ_0.

STEP 1: We start with the following lemma:

Lemma 2. *Let Assumptions A1-A4, and A6 hold. Fix $T \geqslant 1$, $N \geqslant 1$, and let $I(\eta, \lambda) \equiv E\big[\phi^*\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)\big|\mathcal{F}_{t,s_i}\big] - \eta$. Suppose that for some $\bar{\theta} \in \Theta$, $\inf_{(\eta,\lambda')' \in \mathbb{R}^{G+1}} I(\eta, \lambda)$ is attained and $(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta}))$ is optimal with $I(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})) < \infty$ a.s.. If there exists $\mathcal{U}_{\bar{\theta}}$ open in $\Theta \times \mathbb{R}^{G+1}$ such that for every $(\theta', \eta, \lambda')' \in \mathcal{U}_{\bar{\theta}}$ we have: $E\big[\phi^*\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s., $E\big[\phi^{*'}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s., and $E\big[\phi^{*'}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big)|r_{t,s_i}(Y_{t,s_i}, \theta)|\big|\mathcal{F}_{t,s_i}\big] < \infty$ a.s., then the $\mathcal{D}_\phi$-projection $g^*_{t,s_i}(\cdot, \theta)$ of f_{t,s_i} on $\mathcal{SP}_{\bar{\theta},t,i}$ exists, is a.s. P unique and positive for a.e. ω. For every $y \in \mathbb{R}^G$, it is given by:*

$$g^*_{t,s_i}(y, \bar{\theta}) = \phi^{*'}\left(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\right) f_{t,s_i}(y) \text{ a.s..}$$

Next, we show that when $\bar{\theta} = \theta_0$, $\inf_{(\eta,\lambda')' \in \mathbb{R}^{G+1}} I(\eta, \lambda)$ is attained at $(\bar{\eta}_{t,s_i}(\theta_0), \bar{\lambda}_{t,s_i}(\theta_0)')' = 0 \in \mathbb{R}^{G+1}$. For this, we use the strict convexity of ϕ^* (from Lemma 1(ii)) which implies that for any $\upsilon \in \mathbb{R}$, $\phi^*(\upsilon) - \phi^*(0) \geqslant \upsilon \phi^{*'}(0)$. From Lemma 1(v) and Assumption A6(iii) we know that $\phi^{*'}(0) = 1$ and $\phi^*(0) = 0$, so for any $(\eta, \lambda')' \in \mathbb{R}^{G+1}$ we have $I(\eta, \lambda) \geqslant \lambda' E\big[r_{t,s_i}(Y_{t,s_i}, \theta_0)\big|\mathcal{F}_{t,s_i}\big] = 0$ a.s.. So for any $(\eta, \lambda')' \in \mathbb{R}^{G+1}$ it holds that $I(\eta, \lambda) \geqslant I(\bar{\eta}_{t,s_i}(\theta_0), \bar{\lambda}_{t,s_i}(\theta_0)) = 0$ a.s. which shows that $(\bar{\eta}_{t,s_i}(\theta_0), \bar{\lambda}_{t,s_i}(\theta_0)')' = 0 \in \mathbb{R}^{G+1}$ is optimal and that $\inf_{(\eta,\lambda')' \in \mathbb{R}^{G+1}} I(\eta, \lambda)$ is attained.

Now, note that under Assumption A8, we can let $\mathcal{U}_{\theta_0} \equiv \mathcal{U}(\theta_0, \varepsilon_1)$ so the moment conditions of Lemma 2 hold when $\bar{\theta} = \theta_0$.[18] Applying the lemma shows that Theorem 1 holds for $\theta = \theta_0$.

[18] Recall that $\mathcal{U}(\theta_0, \varepsilon_1)$ is an open ball in $\Theta \times \mathbb{R}^{G+1}$ with radius $\varepsilon_1 > 0$ and centered at $(\theta_0', 0, 0')'$.

STEP 2: We now use the set of first order conditions satisfied by $(\bar\eta_{t,s_i}(\theta_0), \bar\lambda_{t,s_i}(\theta_0)')'$. From the proof of Lemma 2, we know that:

$$
\begin{aligned}
0 &= \int_{\mathbb{R}^G} \phi^{*'}\big(\bar\eta_{t,s_i}(\theta_0) + \bar\lambda_{t,s_i}(\theta_0)' r_{t,s}(y,\theta_0)\big) f_{t,s_i}(y)dy - 1 \text{ a.s.} \\
&= \int_{\mathbb{R}^G} \phi^{*'}\big(\bar\eta_{t,s_i}(\theta_0) + \bar\lambda_{t,s_i}(\theta_0)' r_{t,s}(y,\theta_0)\big) r_{t,s}(y,\theta_0) f_{t,s_i}(y)dy \text{ a.s.}
\end{aligned} \tag{28}
$$

Let $\tau \equiv (\eta,\lambda) \in \mathbb{R}^{G+1}$ and $\tau_0 \equiv 0 \in \mathbb{R}^{G+1}$. For any $(\theta,\tau) \in \Theta \times \mathbb{R}^{G+1}$ consider then:

$$\tilde F(\theta,\tau) \equiv \int_{\mathbb{R}^G} F(\theta,\tau,y) f_{t,s_i}(y)dy,$$

where for any $y \in \mathbb{R}^G$ we define:

$$F(\theta,\tau,y) \equiv \begin{pmatrix} \phi^{*'}\big(\eta + \lambda' r_{t,s_i}(y,\theta)\big) - 1 \\ \phi^{*'}\big(\eta + \lambda' r_{t,s_i}(y,\theta)\big) r_{t,s_i}(y,\theta) \end{pmatrix}.$$

Note that under A8, Lemma 2 also shows that $\tau \mapsto \tilde F(\theta,\tau)$ is continuous a.s. on $\mathbb{R}^{G+1} \cap \mathcal{U}_{\theta_0}$. Continuity of $\theta \mapsto \tilde F(\theta,\tau)$ a.s. on $\Theta \cap \mathcal{U}_{\theta_0}$ follows from continuity of ϕ^* (Lemma 1(i)) and $r_{t,s_i}(y,\cdot)$ (Assumption A7), and from Assumption A8(ii,iii) by using the same reasoning as in Lemma 2.

We now establish that $(\theta,\tau) \mapsto \tilde F(\theta,\tau)$ is also continuously differentiable in a neighborhood of (τ_0,θ_0). Under Assumptions A6 and A7, the mapping $(\theta,\tau) \mapsto F(\theta,\tau,y)$ is continuously differentiable on $\Theta \times \mathbb{R}^{G+1}$. Let then $D_\tau F(\theta,\tau,y)' \in L(\mathbb{R}^{G+1},\mathbb{R}^{G+1})$ and $D_\theta F(\theta,\tau,y)' \in L(\mathbb{R}^k,\mathbb{R}^{G+1})$ denote the derivatives of F with respect to τ and θ, respectively. Writing r for $r_{t,s_i}(y,\theta)$ we have:

$$
\begin{aligned}
D_\tau F(\theta,\tau,y) &= \phi^{*''}(\eta+\lambda' r) \begin{pmatrix} 1 & r' \\ r & rr' \end{pmatrix} \\
D_\theta F(\theta,\tau,y) &= \Big(\phi^{*''}(\eta+\lambda' r) D_\theta r \lambda \quad \phi^{*''}(\eta+\lambda' r) D_\theta r \lambda r' + \phi^{*'}(\eta+\lambda' r) D_\theta r\Big)
\end{aligned}
$$

where $D_\theta r' \in L(\mathbb{R}^k,\mathbb{R}^G)$ denotes a partial derivative of $r_{t,s_i}(y,\theta)$ with respect to θ. Using the fact that ϕ^* is convex, we then have:

$$\|D_\tau F(\theta,\tau,y)\| = \phi^{*''}(\eta+\lambda' r)\big(1+|r|^2\big),$$

and

$$\|D_\theta F(\theta,\tau,y)\| \leqslant \phi^{*''}(\eta+\lambda' r)|D_\theta r\lambda| + \phi^{*''}(\eta+\lambda' r)\|D_\theta r\lambda r'\| + \phi^{*'}(\eta+\lambda' r)\|D_\theta r\|.$$

Given the continuity of r, $\phi^{*\prime}$, and $\phi^{*\prime\prime}$, and the moment assumptions in A9, both $\|D_\tau F(\theta,\tau,y)\|$ and $\|D_\theta F(\theta,\tau,y)\|$ are bounded on $\mathcal{U}_{\theta_0} \cap \mathcal{U}(\theta_0,\varepsilon_2)$ by quantities that are integrable with respect to f_{t,s_i}. So by Lebesgue Dominated Convergence Theorem we can exchange limits and integration to get (letting $R \equiv \eta + \lambda' r_{t,s_i}(Y_{t,s_i},\theta)$), with probability one:

$$\begin{aligned} &D_\tau \tilde{F}(\theta,\tau) \\ &\quad= \begin{pmatrix} E[\phi^{*''}(R)|\mathcal{F}_{t,s_i}] & E[\phi^{*''}(R) r_{t,s_i}(Y_{t,s_i},\theta)|\mathcal{F}_{t,s_i}]' \\ E[\phi^{*''}(R) r_{t,s_i}(Y_{t,s_i},\theta)|\mathcal{F}_{t,s_i}] & E\big[\phi^{*''}(R) r_{t,s_i}(Y_{t,s_i},\theta) r_{t,s_i}(Y_{t,s_i},\theta)'|\mathcal{F}_{t,s_i}\big] \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} &D_\theta \tilde{F}(\theta,\tau) \\ &= \Big(E\big[\phi^{*''}(R) D_\theta r_{t,s_i}(Y_{t,s_i},\theta)\lambda\big|\mathcal{F}_{t,s_i}\big] \quad E\big[\phi^{*''}(R) D_\theta r_{t,s_i}(Y_{t,s_i},\theta)\lambda r_{t,s_i}(Y_{t,s_i},\theta)' + \phi^{*'}(R) D_\theta r_{t,s_i}(Y_{t,s_i},\theta)\big|\mathcal{F}_{t,s_i}\big]\Big) \end{aligned}$$

for all $(\theta,\tau) \in \mathcal{U}_{\theta_0} \cap \mathcal{U}(\theta_0,\varepsilon_2)$. Same assumptions suffice to show that $(\theta,\tau) \mapsto D_\tau \tilde{F}(\theta,\tau)$ and $(\theta,\tau) \mapsto D_\theta \tilde{F}(\theta,\tau)$ are continuous on $\mathcal{U}_{\theta_0} \cap \mathcal{U}(\theta_0,\varepsilon_2)$, following a reasoning similar to that in the proof of Lemma 2. In particular, under Assumption A3 we have:

$$\begin{aligned} D_\tau \tilde{F}(\theta_0,\tau_0) &= \begin{pmatrix} 1 & 0 \\ 0 & E\big[r_{t,s_i}(Y_{t,s_i},\theta_0) r_{t,s_i}(Y_{t,s_i},\theta_0)'|\mathcal{F}_{t,s_i}\big] \end{pmatrix} \text{ a.s.} \\ D_\theta \tilde{F}(\theta_0,\tau_0) &= \begin{pmatrix} 0 & E[D_\theta r_{t,s_i}(Y_{t,s_i},\theta_0)|\mathcal{F}_{t,s_i}] \end{pmatrix} \text{ a.s..} \end{aligned}$$

Finally, we invoke Implicit Function Theorem for (θ,τ) in a neighborhood of (θ_0,τ_0), which by Equation (28) are known to solve $\tilde{F}(\theta_0,\tau_0) = 0$ a.s.. Under Assumption A10, $D_\tau \tilde{F}(\theta_0,\tau_0)$ is invertible a.s.. Then Implicit Function Theorem (e.g. Theorem 9.28 in Rudin (1976)) applies and there exists $\mathcal{B}(\theta_0,\varepsilon)$ in which to any $\theta \in \mathcal{B}(\theta_0,\varepsilon) \subset \Theta$ there corresponds a unique τ such that:

$$(\theta,\tau) \in \mathcal{U}_{\theta_0} \cap \mathcal{U}(\theta_0,\varepsilon_2) \text{ and } \tilde{F}(\tau(\theta),\theta) = 0 \text{ a.s.}$$

In particular, for each $\bar{\theta} \in \mathcal{B}(\theta_0, \varepsilon)$, there exists a unique $(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})) \equiv \tau(\bar{\theta})$ which satisfies the conditions in Equation (29); so $(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta}))$ minimizes $I(\eta, \lambda)$ when $\theta = \bar{\theta}$. Given that $(\bar{\theta}', \bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})')' \in \mathcal{U}_{\theta_0} \cap \mathcal{U}(\theta_0, \varepsilon_2)$, the moment conditions of Lemma 2 are satisfied and we can apply its results to show that the $\mathcal{D}_\phi$-projection $g^*_{t,s_i}(y, \bar{\theta})$ of f_{t,s_i} on $\mathcal{SP}_{\bar{\theta},t,s_i}$ exists, is a.s. P unique and positive for a.e. ω, and given by:

$$g^*_{t,s_i}(y, \bar{\theta}) = \phi^{*'}\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big) f_{t,s_i}(y) \text{ a.s..}$$

for every $y \in \mathbb{R}^G$. In addition, the mapping $\bar{\theta} \mapsto \bar{\tau} = \tau(\bar{\theta})$ is continuously differentiable on $\mathcal{B}(\theta_0, \varepsilon)$ with derivative $D_{\bar{\theta}}\tau(\bar{\theta})' \in L(\mathbb{R}^k, \mathbb{R}^{G+1})$, and when $\bar{\theta} = \theta_0$ (and $\bar{\tau} = \tau_0$) we have: $D_\theta\tau(\theta_0) = -D_\theta\tilde{F}(\tau_0, \theta_0)\big[D_\tau\tilde{F}(\tau_0, \theta_0)\big]^{-1}$. In particular, under Assumption A3 we have:

$$\begin{aligned} D_\tau\tilde{F}(\tau_0, \theta_0) &= \begin{pmatrix} 1 & 0 \\ 0 & E\big[r_{t,s_i}(Y_{t,s_i}, \theta_0) r_{t,s_i}(Y_{t,s_i}, \theta_0)' | \mathcal{F}_{t,s_i}\big] \end{pmatrix} \text{ a.s.} \\ D_\theta\tilde{F}(\tau_0, \theta_0) &= \begin{pmatrix} 0 & E[D_\theta r_{t,s_i}(Y_{t,s_i}, \theta_0) | \mathcal{F}_{t,s_i}] \end{pmatrix} \text{ a.s.} \end{aligned}$$

which shows that

$$D_\theta\tau(\theta_0) = \Big(0 \quad -E[D_\theta r_{t,s_i}(Y_{t,s_i}, \theta_0)|\mathcal{F}_{t,s_i}]\big\{E\big[r_{t,s_i}(Y_{t,s_i}, \theta_0) r_{t,s_i}(Y_{t,s_i}, \theta_0)'|\mathcal{F}_{t,s_i}\big]\big\}^{-1}\Big).$$

This completes the proof of Theorem 1 and its Corollaries 2 and 3. □

Proof of Lemma 2. The proof is done in two steps. We first show that $g^*_{t,s_i}(y, \bar{\theta})$ defined in Lemma 2 is in $\mathcal{SP}_{\bar{\theta},t,s_i}$. Then, we show that it is optimal.

STEP 1: Given $\bar{\theta}$, we have that $I(\eta, \lambda) < \infty$ a.s. for any $(\eta, \lambda')' \in \mathbb{R}^{G+1} \cap \mathcal{U}_{\bar{\theta}}$ which is open. Hence, $(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta}))$ is an interior optimum, and we have that $D_\eta I(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})) = 0$ a.s., and $D_\lambda I(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})) = 0$ a.s., where $D_\eta I(\eta, \lambda)' \in L(\mathbb{R}, \mathbb{R})$ and $D_\lambda I(\eta, \lambda)' \in L(\mathbb{R}^G, \mathbb{R})$ denote the partial derivatives of I with respect to η and λ. We first use Lebesgue Dominated Convergence Theorem to be able to take the limit into the expectation in:

$$\begin{aligned} &D_\eta I(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})) = \\ &\quad \lim_{h \to 0} E\Big[\frac{\phi^*\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + h + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s_i}(Y_{t,s_i}, \bar{\theta})\big) - \phi^*\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s_i}(Y_{t,s_i}, \bar{\theta})\big)}{h}\Big|\mathcal{F}_{t,s_i}\Big] - 1 \text{ a.s..} \end{aligned}$$

Under Assumption A6, Lemma 1 applies and ϕ^* is in $\mathcal{C}^2(\mathbb{R},\mathbb{R})$ so by mean value theorem:

$$\frac{\phi^*\big(\bar\eta_{t,s_i}(\bar\theta)+h+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)-\phi^*\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)}{h}=$$
$$\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\dot h+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)\text{ a.s.},$$

with $\dot h\in(\min\{0,h\},\max\{0,h\})$. Given that $\phi^{*\prime}$ is positive and strictly increasing on $\mathbb{R}$ (see Lemma 1(iv)), we have:

$$0<\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\dot h+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)\leqslant\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\max\{0,h\}+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)\text{ a.s..}$$

Now, for $h\in\mathbb{R}$ such that $(\bar\theta',\bar\eta_{t,s_i}(\bar\theta)+h,\bar\lambda_{t,s_i}(\bar\theta)')'\in\mathcal{U}_{\bar\theta}$, the upper bound above is integrable with respect to f_{t,s_i}; we can therefore exchange limit and expectation to get:

$$D_\eta I(\bar\eta_{t,s_i}(\bar\theta),\bar\lambda_{t,s_i}(\bar\theta))=E\Big[\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)\Big|\mathcal{F}_{t,s_i}\Big]-1\text{ a.s..}$$

The same reasoning shows that for any $(\theta',\eta,h,\lambda')'\in\Theta\times\mathbb{R}^{G+2}$ such that $(\theta',\eta,\lambda')'\in\mathcal{U}_{\bar\theta}$ and $(\theta',\eta+h,\lambda')'\in\mathcal{U}_{\bar\theta}$, we have:

$$\lim_{h\to0}E\big[\phi^{*\prime}\big(\eta+h+\lambda'r_{t,s_i}(Y_{t,s_i},\theta)\big)\big|\mathcal{F}_{t,s_i}\big]=E\big[\phi^{*\prime}\big(\eta+\lambda'r_{t,s_i}(Y_{t,s_i},\theta)\big)\big|\mathcal{F}_{t,s_i}\big]\text{ a.s.},$$

so that $\eta\mapsto E\big[\phi^{*\prime}\big(\eta+\lambda'r_{t,s_i}(Y_{t,s_i},\theta)\big)\big|\mathcal{F}_{t,s_i}\big]$ is continuous a.s. on $\mathbb{R}\cap\mathcal{U}_{\bar\theta}$.

Similarly, fix any $1\leqslant j\leqslant G$ and consider the partial derivative of $I(\eta,\lambda)$ with respect to λ_j, when evaluated at $(\bar\eta_{t,s_i}(\bar\theta),\bar\lambda_{t,s_i}(\bar\theta))$. We have:

$$\frac{\phi^*\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)+hr^j_{t,s_i}(Y_{t,s_i},\bar\theta)\big)-\phi^*\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big)}{h}$$
$$=\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)+\dot hr^j_{t,s_i}(Y_{t,s_i},\bar\theta)\big)r^j_{t,s_i}(Y_{t,s_i},\bar\theta)\text{ a.s.},$$

where r^j denotes the jth component of r, and $\dot h\in(\min\{0,h\},\max\{0,h\})$. Now, using again the convexity of ϕ^* we have, with probability one:

$$\Big|\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)+\dot hr^j_{t,s_i}(Y_{t,s_i},\bar\theta)\big)r^j_{t,s_i}(Y_{t,s_i},\bar\theta)\Big|\leqslant$$
$$\max\Big\{\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)\big),\phi^{*\prime}\big(\bar\eta_{t,s_i}(\bar\theta)+\bar\lambda_{t,s_i}(\bar\theta)'r_{t,s_i}(Y_{t,s_i},\bar\theta)+hr^j_{t,s_i}(Y_{t,s_i},\bar\theta)\big)\Big\}\big|r^j_{t,s_i}(Y_{t,s_i},\bar\theta)\big|$$

Both terms of the right hand side of the above inequality are integrable with respect to f_{t,s_i}, so using again Lebesgue's Dominated Convergence theorem, we get:

$$D_{\lambda_j} I(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta})) = E\Big[\phi^{*'}\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s_i}(Y_{t,s_i}, \bar{\theta})\big) r^j_{t,s_i}(Y_{t,s_i}, \bar{\theta}) \Big| \mathcal{F}_{t,s_i}\Big] \text{ a.s..}$$

Same reasoning as previously shows that moreover, for any $(\theta', \eta, \lambda')' \in \mathcal{U}_{\bar{\theta}}$, we have $\lambda_j \mapsto E\big[\phi^{*\prime}\big(\eta + \lambda' r_{t,s_i}(Y_{t,s_i}, \theta)\big) r^j_{t,s_i}(Y_{t,s_i}, \bar{\theta}) \big| \mathcal{F}_{t,s_i}\big]$ continuous a.s. on $\mathbb{R} \cap \mathcal{U}_{\bar{\theta}}$.

In particular, the first order conditions satisfied by $(\bar{\eta}_{t,s_i}(\bar{\theta}), \bar{\lambda}_{t,s_i}(\bar{\theta}))$ can then be written as:

$$\begin{aligned} 0 &= \int_{\mathbb{R}^G} \phi^{*'}\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big) f_{t,s_i}(y) dy - 1 \text{ a.s.} \\ &= \int_{\mathbb{R}^G} \phi^{*'}\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big) r_{t,s}(y, \bar{\theta}) f_{t,s_i}(y) dy \text{ a.s..} \end{aligned} \tag{29}$$

Combined with $\phi^{*'} > 0$ on $\mathbb{R}$ from Lemma 1(iv), and $f_{t,s_i}(y) > 0$ a.s. for every $y \in \mathbb{R}^G$ from Assumption A4, the two equalities in Equation (29) show that $g^*_{t,s_i}(\cdot, \bar{\theta})$ is a feasible element of $\mathcal{SP}_{\bar{\theta},t,s_i}$.

STEP 2: We now show that $g^*_{t,s_i}(\cdot, \bar{\theta})$ is indeed optimal. Let $\pi_{\bar{\theta},t,s_i}$ be any other probability density belonging to $\mathcal{SP}_{\bar{\theta},t,s_i}$. As consequence of Assumption A6, we have that for all $(\upsilon, u) \in \mathbb{R}^2$ (see Hiriart-Urruty and Lemarechal (1993)):

$$\phi^*(\upsilon) = \upsilon(\phi')^{-1}(\upsilon) - \phi\left((\phi')^{-1}(\upsilon)\right) \geqslant \upsilon u - \phi(u).$$

When evaluated at $u \equiv \pi_{\bar{\theta},t,s_i}(y)/f_{t,s_i}(y)$ and $\upsilon \equiv \bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})$, the above inequality becomes:

$$\begin{aligned} &\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big)(\phi')^{-1}\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big) f_{t,s_i}(y) \\ &\qquad - \phi\left((\phi')^{-1}\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big)\right) f_{t,s_i}(y) \\ &\qquad\quad \geqslant \pi_{\bar{\theta},t,s_i}(y)\big(\bar{\eta}_{t,s_i}(\bar{\theta}) + \bar{\lambda}_{t,s_i}(\bar{\theta})' r_{t,s}(y, \bar{\theta})\big) - \phi\Big(\frac{\pi_{\bar{\theta},t,s_i}(y)}{f_{t,s_i}(y)}\Big) f_{t,s_i}(y) \text{ a.s..} \end{aligned}$$

Integrating over $\mathbb{R}^G$, using Equation (29) and feasibility of the probability density $\pi_{\bar\theta,t,s_i}$ then gives:

$$\begin{aligned}\mathcal{D}_\phi(g^*_{t,s_i}(\cdot,\bar\theta), f_{t,s_i}) &= \int_{\mathbb{R}^G} \phi\left(\phi^{*'}\left(\bar\eta_{t,s_i}(\bar\theta) + \bar\lambda_{t,s_i}(\bar\theta)' r_{t,s}(y,\bar\theta)\right)\right) f_{t,s_i}(y)dy \\ &\leqslant \int_{\mathbb{R}^G} \phi\Big(\frac{\pi_{\bar\theta,t,s_i}(y)}{f_{t,s_i}(y)}\Big) f_{t,s_i}(y)dy = \mathcal{D}_\phi(\pi_{\bar\theta,t,s_i}, f_{t,s_i}) \text{ a.s.,}\end{aligned}$$

so $g^*_{t,s_i}(\cdot,\bar\theta)$ is optimal. This completes the proof of Lemma 2. □

Proof of Theorem 4. We start the proof with the following lemma:

Lemma 3. *Suppose Assumptions A11, A13 hold, and Z is α-mixing with mixing coefficients $\alpha_{k,l}(m)$. Then, for every $\theta \in \Theta'$, $\{G_{t,s_i}(\theta), (t,i) \in \mathbb{N}^2\}$ is α-mixing with mixing coefficients $\alpha_{k,l}(m)$.*

Combining Assumption A12(a), Assumption A13 and Lemma 3, it then follows that for all $\theta \in \Theta'$, $G_{t,s_i}(\theta)$ is α-mixing with mixing coefficient satisfying $\sum_{m=1}^\infty m\,\alpha_{1,1}(m) < \infty$. By this and Assumption A14(i) we have, pointwise on Θ', $(TN)^{-1}\sum_{t=1}^T\sum_{i=1}^N G_{t,s_i}(\theta) - E\big[G_{t,s_i}(\theta)\big] \xrightarrow{p} 0$ (see Jenish and Prucha, 2007, Theorem 3). The Lipschitz condition of Assumption A14(ii) suffices for $G_{t,s_i}(\theta)$ to be L_0 stochastically equicontinuous on Θ' (see Jenish and Prucha, 2007, Proposition 1). Thus, we can apply Jenish and Prucha (2007, Theorem 2) to get that, as $(T,N)\to\infty$, $\sup_{\theta\in\Theta'}|(TN)^{-1}\sum_{t=1}^T\sum_{i=1}^N G_{t,s_i}(\theta) - E\big[G_{t,s_i}(\theta)\big]| \xrightarrow{p} 0$. Therefore, consistency of θ_{TN} follows by standard asymptotic arguments (e.g. Gallant and White, 1988). □

Proof of Lemma 3. From Assumption A11, $r_{t,s}(Y_{t,s},\theta)$ is continuous with respect to all its arguments which implies that $r_{t,s_i}(Y_{t,s_i},\theta)$ is a $\mathcal{F}_{t,s_i}$-measurable function of $(Y_{t,s_i},\dots,Y_{t-\tau,s_i},X_{t,s_i},\dots,X_{t-\tau,s_i})$ with τ finite. Hence, when Z is α-mixing with mixing coefficients $\alpha_{k,l}(m)$ satisfying A12, so is the random field $\{r_{t,s_i}(Y_{t,s_i},\theta),(t,i)\in\mathbb{N}^2\}$. Assumption A13 guarantees that the same holds for the fields $\{\eta_{t,s_i}(\theta),(t,i)\in\mathbb{N}^2\}$ and $\{\lambda_{t,s_i}(\theta),(t,i)\in\mathbb{N}^2\}$, thereby yielding the desired result. □

Proof of Theorem 5. Consider the mean value expansion of the first order conditions defining θ_{TN}:

$$(30)\quad 0 = \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} D_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \theta_0) + \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} H_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \dot{\theta}_{TN}) \sqrt{TN}(\theta_{TN} - \theta_0),$$

where $H_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \cdot)$ is evaluated at the mean value $\dot{\theta}_{TN} = \{\theta^1_{TN}, \ldots, \theta^K_{TN}\}$ lying on the segment connecting θ_{TN} and θ_0. Assumption A15, the α-mixing conditions of Assumption A12, and Lemma 3 imply Jenish and Prucha's (2007) ULLN for the random field $\{H_\theta \ln g^*_{t,s_i}(Y_{t,s_i}, \theta), (t,i) \in \mathbb{N}^2\}$. By consistency of θ_{TN}, $\dot{\theta}^j_{TN} \xrightarrow{p} \theta^j_0$ we have:

$$(31)\quad \sqrt{TN}(\theta_{TN} - \theta_0) = -H^{-1} \left\{ \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} D_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \theta_0) \right\}$$
$$+ \left[\frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} H_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \dot{\theta}_{TN}) - H \right] \left\{ \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} D_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \theta_0) \right\}$$

By Lemma 3 the score is α-mixing with mixing coefficient satisfying A12. By Assumption A15(i) the first term on the right hand side of rightmost term obeys the CLT for random fields (see Jenish and Prucha, 2007, Theorem 1):

$$(32)\qquad \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} \upsilon' V^{-1/2} D_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \theta_0) \xrightarrow{d} N(0,1).$$

Triangular inequality and uniform convergence of the Hessian on Θ' implied by Assumption A15(ii) gives

$$\left| \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} H_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \dot{\theta}_{TN}) - H \right| = o_p(1) O_p(T^{-1/2} N^{-1/2}) = o_p(1).$$

Therefore,

$$(33)\qquad \sqrt{TN}(\theta_{TN} - \theta_0) = -H^{-1} \left\{ \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} D_\theta \ln g^*_{t,s_i}(y_{t,s_i}, \theta_0) \right\} + o_p(1).$$

Standard multivariate asymptotic normality arguments (see White (2001), Corollary 4.24) give: $\sqrt{TN}(\theta_{TN} - \theta_0) \xrightarrow{d} N\big(0, H^{-1} V H^{-1}\big)$. In virtue of Assumption A3 and equation (13), the moment functions $r_{t,s_i}(Y_{t,s_i},\theta_0)$ are uncorrelated for every $(t,i)\in\mathbb{N}^2$ and, thus, the following simplification occurs:

$$V=\lim_{T,N\to\infty}\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}E\Bigg(D_\theta\lambda_{t,s_i}(\theta_0)'r_{t,s_i}(Y_{t,s_i},\theta_0)r_{t,s_i}(Y_{t,s_i},\theta_0)'D_\theta\lambda_{t,s_i}(\theta_0)\Bigg)$$

Hence, application of the law of iterated expectations gives that $V=H$, from which it follows that

$$\sqrt{TN}(\theta_{TN}-\theta_0)\xrightarrow{d}N(0,H^{-1})$$

as required. $\square$

References

Ahn, S. C. and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

Ali, S. M. and S. D. Silvey (1966): "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society Ser. B*, 28, 131–142.

Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

Arellano, M. and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, 3231–3296.

Baltagi, B. H. and H. M. Pesaran (2007): "Heterogeneity and Cross Section Dependence in Panel Data Models: Theory and Applications," *Journal of Applied Econometrics*, 22, 229–232.

Bickel, P. J. (1982): "On Adaptive Estimation," *Annals of Statistics*, 10, 647–671.

Bolthausen, E. (1982): "On the Central Limit Theorem for Stationary Mixing Random Fields," *The Annals of Probability*, 10, 1047–1050.

Borovkov, A. A. and B. A. Rogozin (1965): "On the Multi-Dimensional Central Limit Theorem," *Theory of Probability and its Applications*, 10, 55–62.

Bradley, R. C. (1989): "A Caution on Mixing Conditions for Random Fields," *Statistics & Probability Letters*, 8, 489–491.

Bulinsky, A. V. and I. Zhurbenko (1976): "Central Limit Theorem for Additive Random Functions," *Teor. Veroyatnost. i Primenen*, 21, 707–717, (Russian). English transl. in: Theory Probab. Appl. 21, 687-697.

Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels (2007): "Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions," *Journal of Econometrics*, 140, 529–573.

Carrasco, M. and J.-P. Florens (2004): "On the Asymptotic Efficiency of GMM," Manuscript, University of Montreal AND University of Toulouse I.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1992): "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business & Economic Statistics*, 10, 20–26.

Cramér, H. (1938): "Sur un Nouveau Théorème-Limite de la Théorie Des Probabilités," *Actualités Scientifiques et Industrielles*, 736.

Csiszar, I. (1967): "Information-type measures of the difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, 299–318.

——— (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, 3, 146–58.

——— (1995): "Generalized Projections for non-negative functions," *Acta Math. Hungar.*, 68, 161–185.

Davidson, J. (1994): *Stochastic Limit Theory*, Oxford University Press.

Dobruschin, P. L. (1968): "The Description of a Random Field by Means of Conditional Probabilities and Conditions of its Regularity," *Theory of Probability and its Applications*, 13, 197–224.

Durrett, R. (1995): *Probability: Theory and Examples*, Duxbury Press.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Föllmer, H. (1988): "Random Fields and Diffusion Processes," in *École D'Été de Probabilités de Saint-Flour XV-XVII, 1985-87*, Berlin / Heidelberg: Springer, 101–203.

Gallant, R. A. and H. White (1988): *A unified theory of estimation and inference for nonlinear dynamic models*, New York: Basil Blackwell.

Hahn, J. and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica*, 70, 1639–1657.

——— (2004): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," Manuscript, UCLA and UC Davis.

Hayashi, F. (1992): "[On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous]: Comment," *Journal of Business & Economic Statistics*, 10, 15–17.

Hiriart-Urruty, J.-B. and C. Lemarechal (1993): *Convex Analysis and Minimization Algorithms I: Fundamentals (Grundlehren Der Mathematischen Wissenschaften)*, Springer.

Ivanov, A. and N. Leonenko (1986): *Statistical Analysis of Random Fields*, Kluwer Academic Publishers, translated from Russian. Translation Edited by S. Kotz.

Jenish, N. and I. R. P. Prucha (2007): "Central Limit Theorems and uniform Laws of Large Numbers for Arrays of Random Fields," University of Maryland, College Park.

Keane, M. P. and D. E. Runkle (1992): "On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business & Economic Statistics*, 10, 1–9.

Kitamura, Y. and M. Stutzer (1997): "An information-theoretic alternative to generalized method of moments estimation," *Econometrica*, 65, 861–74.

Koopmans, T. C. (1950): "When is an Equation System Complete for Statistical Purposes?" in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, John Wiley & Sons, Inc., 393–409.

Kullback, S. and M. A. Khairat (1966): "A Note on Minimum Discrimination Information," *The Annals of Mathematical Statistics*, 37, 279–280.

Palais, R. S. (1959): "Natural Operations on Differential Forms," *Transactions of the American Mathematical Society*, 92, 125–141.

Park, B. U., R. C. Sickles, and L. Simar (2007): "Semiparametric Efficient Estimation of Dynamic Panel Data Models," *Journal of Econometrics*, 136, 281–301.

Phillips, P. C. B. and H. R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press.

Rosenblatt, M. (1986): "Parameter Estimation for Finite-Parameter Stationary Random Fields," *Journal of Applied Probability*, 23, 311–318.

Rudin, W. (1976): *Principles of Mathematical Analysis*, McGraw-Hill.

Schmidt, P., S. C. Ahn, and D. Wyhowski (1992): "[On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous]: Comment," *Journal of Business & Economic Statistics*, 10, 10–14.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, vol. 1, 187–196.

Teboulle, M. and I. Vajda (1993): "Convergence of Best ϕ-Entropy Estimates," *IEEE Transactions on Information Theory*, 39, 297–301.

White, H. (2001): *Asymptotic Theory for Econometricians*, San Diego: Academic Press.

Wooldridge, J. M. (2005): "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

Appendix A. Overview of Random Fields

In this Appendix, we give a brief overview of random fields. Let $Z \equiv \{Z_s, s \in \mathbb{N}^2\}$ be a collection of random vectors $Z_s \in \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{Z}, P)$. Let $(\mathbb{R}^d)^{\mathbb{N}^2}$ denote the countable product space; then $Z : \Omega \to (\mathbb{R}^d)^{\mathbb{N}^2}$. For any $S \subset \mathbb{N}^2$, we introduce the σ-algebra $\mathcal{Z}_S \equiv \sigma(Z_s : s \in S)$. Note that $\mathcal{Z} = \mathcal{Z}_{\mathbb{N}^2}$. We will usually identify Z with its distribution which is a probability measure μ on $(\mathbb{R}^d)^{\mathbb{N}^2}$. In the approach of Dobruschin (1968), μ is called *a random field* (see also Föllmer (1988)). Here, we shall use the same nomenclature—random field—for both Z and its distribution μ.

For any $V \subset \mathbb{N}^2$ we can choose a regular conditional distribution of μ with respect to $\mathcal{Z}_{V^c}$ (where $V^c = \mathbb{N}^2 \backslash V$ is the usual complement of V), $\mu_V : \Omega \times \mathcal{Z}_{V^c} \to \mathbb{R}_+$ such that: (i) for every $A \in \mathcal{Z}_{V^c}$, $\omega \mapsto \mu_V(\omega, A)$ is a version of a conditional distribution of μ given $\mathcal{Z}_{V^c}$, i.e. for any $\mathcal{Z}$-measurable function $\varphi \geqslant 0$ we have:

$$E_\mu[\varphi|\mathcal{Z}_{V^c}](\omega) = \int \varphi(A)\mu_V(\omega, dA) \equiv (\mu_V\varphi)(\omega); \tag{34}$$

and, (ii) for a.e. ω, $A \mapsto \mu_V(\omega, A)$ is a probability measure on $(\Omega, \mathcal{Z}_{V^c})$ (Föllmer, 1988). These conditional distributions are consistent in the usual sense: since $\mathcal{Z}_{W^c} \subseteq \mathcal{Z}_{V^c}$ for $W \supseteq V$, we have:

$$(\mu_W\mu_V\varphi)(\omega) = E_\mu\big[E_\mu[\varphi|\mathcal{Z}_{V^c}]\big|\mathcal{Z}_{W^c}\big](\omega) = E_\mu[\varphi|\mathcal{Z}_{W^c}](\omega) = (\mu_W\varphi)(\omega) \text{ a.s.} \tag{35}$$

for any $\mathcal{Z}$-measurable $\varphi \geqslant 0$.

We are now going to prescribe the local conditional behavior of a random field by fixing a system of conditional distributions μ_V for the finite subsets $V \subseteq \mathbb{N}^2$. Note that these conditional distributions are required to be consistent in a strict sense, i.e. without the intervention of null sets.

Definition 5 (**Gibbs Measure**). For each finite $V \subseteq \mathbb{N}^2$, let $\mu_V : \Omega \times \mathcal{Z}_{V^c} \to \mathbb{R}_+$ be such that for a.e. ω, $A \mapsto \mu_V(\omega, A)$ is a probability measure on $(\Omega, \mathcal{Z}_{V^c})$. The collection (μ_V) is called *a local specification* if $\mu_W\mu_V = \mu_W$ for $V \subseteq W$. A random field μ is called *a Gibbs* *measure* with respect to the local specification (μ_V) if, for any finite V, μ_V is a conditional distribution of μ with respect to $\mathcal{Z}_{V^c}$ in the sense of Equation (34).

In what follows, for a given local specification (μ_V) we denote by $G(\mu)$ the corresponding class of Gibbs measures. It is worth pointing out that $G(\mu)$ need not be nonempty nor a singleton. Dobruschin (1968) investigates the problems of existence and uniqueness of a random field μ with a given system of conditional distributions (μ_V).

We now discuss some important properties of the random fields.

Definition 6 (**Spatial Homogeneity**). Let $i \in \mathbb{N}^2$ and let $T_i : \Omega \to \Omega$ be the shift map defined by $(T_i\omega)(k) = \omega(i + k)$. A local specification is called *spatially homogeneous* if $\mu_V(T_i\omega, \cdot) = \mu_{V+i}(\omega, \cdot) \circ T_i$, i.e. $(\mu_V\varphi) \circ T_i = \mu_{V+i}(\varphi \circ T_i)$ for finite $V \subseteq \mathbb{N}^2$.

In this case, we denote by $S_\Omega \equiv \{\mu : \mu \circ T_i = \mu \text{ for any } i \in \mathbb{N}^2\}$ the class of all spatially homogeneous random fields. In the literature, the name "spatial homogeneity" seems to be reserved for the measure μ (see e.g. Föllmer (1988)). An equivalent property of Z is referred to as "strict stationarity" (Rosenblatt, 1986; Bradley, 1989).

Definition 7 (**Strict Stationarity**). The random field Z is strictly stationary, if the associated measure μ is spatially homogeneous, i.e. for any $A \subseteq \mathcal{Z}$, $P(Z \in T_iA) = \mu(T_iA) = \mu(A) = P(Z \in A)$ where $i \in \mathbb{N}^2$.

We now explore some dependence properties of the random field Z that extend the usual mixing conditions on random sequences. It is important to note that for strictly stationary random fields such as Z, apparently natural versions of the ϕ-mixing condition turn out to be extremely restrictive. For example, Bradley (1989) showed that versions of ϕ-mixing or even absolute regularity are in fact equivalent to corresponding versions of m-dependence for random sequences. By Rosenblatt (1986), this does not apply to corresponding versions of the α-mixing or ρ-mixing conditions. Thus, we limit our attention to strictly stationary random fields Z that are α-mixing (or "strong" mixing). Rosenblatt (1986) gives the following definition.

Definition 8 (**Strong Mixing**)**.** Let $V \subseteq \mathbb{N}^2$ and $W \subseteq \mathbb{N}^2$ be two sets of indices, and let $d(V,W)$ be the Euclidean distance between them, i.e. $d(V,W) \equiv \inf_{i \in V, j \in W} \|i-j\|$ with $\|i-j\| \equiv \max\{|i_1 - j_1|, |i_2 - j_2|\}$ for any $i = (i_1, i_2) \in \mathbb{N}^2$ and $j = (j_1, j_2) \in \mathbb{N}^2$. Consider $Z_V \equiv (Z_i, i \in V)$ and $Z_W \equiv (Z_j, j \in W)$, and the associated σ-algebras $\mathcal{Z}_V$ and $\mathcal{Z}_W$. The random field Z is said to be α-mixing (or strong mixing) if

$$\alpha(\mathcal{Z}_V, \mathcal{Z}_W) \equiv \sup_{A \in \mathcal{Z}_V,\, B \in \mathcal{Z}_W} [P(Z_V \in A \cap Z_W \in B) - P(Z_V \in A)P(Z_W \in B)] \leqslant \varphi(d(V,W))$$

with φ a function such that $\varphi(d) \to 0$ as $d \to \infty$.

For nonempty sets $V \subseteq \mathbb{N}^2$ and $W \subseteq \mathbb{N}^2$ that are *disjoint*, we use the abbreviation $\alpha(V,W) \equiv \alpha(\mathcal{Z}_V, \mathcal{Z}_W)$. For any $n \in \mathbb{N}$ and $(k,l) \in \mathbb{N} \cup \{\infty\}$, the mixing coefficients for the random field Z are defined as in Bolthausen (1982):

$$\alpha_{k,l}(n) \equiv \sup\{\alpha(V,W) : |V| \leqslant k, |W| \leqslant l, d(V,W) \geqslant n\} \tag{36}$$

where $|V|$ and $|W|$ denote the cardinalities of the sets V and W, respectively.

Let V_n be any subset of $\mathbb{N}^2$. We denote by $|V_n|$ its cardinality, and we let ∂V_n be the boundary of this set, i.e. $\partial V_n \equiv \{i \in V_n : \exists j \notin V_n \text{ such that } |i-j| = 1\}$. Throughout, $(V_n, n \in \mathbb{N})$ is a sequence of finite subsets of $\mathbb{N}^2$, satisfying:

$$\lim_{n \to \infty} |V_n| = \infty \text{ and } \lim_{n \to \infty} |V_n|^{-1} |\partial V_n| = 0 \tag{37}$$