Dynamic Spatial Discrete Choice Using One Step GMM: An Application to Mine Operating Decisions*

Joris Pinkse†, Margaret Slade‡, and Lihong Shen§

October 2005

Abstract

We show that the one–step ('continuous updating') GMM estimator is consistent and asymptotically normal under weak conditions that allow for generic spatial and time series dependence. We use the new procedure to estimate a dynamic spatial discrete–choice model with fixed effects that enables us to study operating decisions for mines in a real–options context. We find that the data are more supportive of a mean/variance–utility model than of a real–options model.

*We thank Tim Conley and seminar participants at Cemmap, the universities of Tilburg and Warwick, and the Tinbergen Institute for their valuable comments. Margaret Slade would like to acknowledge financial support from the ESRC and the Leverhulme Foundation.

†Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park PA 16802, joris@psu.edu

‡Department of Economics, The University of Warwick, Coventry CV4 7AL, United Kingdom, m.slade@warwick.ac.uk

§Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park PA 16802, lshen@psu.edu

1 Introduction

In this paper, we prove the consistency and asymptotic normality of the one–step GMM or continuous updating (CU) estimator of Hansen, Heaton and Yaron (1996) under assumptions that are more plausible in many economic applications than those that are made in the existing spatial literature. We use the CU estimator to estimate a nonlinear dynamic spatial discrete–choice panel–data model with fixed effects that allows us to study mine operating decisions in a real–options context.

The theoretical literature has thus far, implicitly or explicitly, treated spatial dependence as a simple multivariate extension of time–series dependence.[1] Observations are typically regarded as draws from a stationary underlying process (SUP). In many interesting economic applications, however, including ours, spatial dependence can be nonstationary. Indeed, in addition to distance, both location and the number of observations can affect dependence, and both can be endogenous.[2] To establish our results, we make use of a new central limit theorem (CLT) for spatially dependent processes (Pinkse, Shen, and Slade 2005) that covers those eventualities.

Conley (1999) established generic convergence results for the standard two–step GMM estimator. We establish the asymptotic properties of a different GMM–type estimator — the CU estimator — which is a member of the class of generalized empirical likelihood (GEL) estimators. GEL estimators, like standard GMM estimators, use moment conditions. Moreover, in exactly identified models, the two classes of estimators are identical. In overidentified models, however, even though their asymptotic distributions are identical, the statistical properties of GEL estimators tend (or can be made) to be superior in small and moderate–size samples (see, e.g., Newey and Smith 2003).

Our spatial CU procedure is formally stated in a spatial cross–sections context. The results, however, carry over to two different types of panel–data models. In the first model, the number of 'products' (mines in our application) increases while the number of time periods is assumed fixed. This allows for a completely general time–series–dependence structure. With a fixed number of time periods, a panel–data model is equivalent to a (spatial) cross–sections model with a larger number of moment conditions. Furthermore, it is comparatively easy to find suitable instruments for that model. An alternative possibility is that both the time–series and cross–sections dimensions grow. In that case, if one assumes weak dependence in the time–series dimension, the temporal dimension can be treated as an additional spatial dimension.

Since we have potentially nonstationary data, we introduce a Newey–West (1987) style covariance–matrix estimator for nonstationary spatial data. That estimator simplifies to the Newey–West estimator in the case of a stationary time series.

It is common for panel–data sets to have a large number of cross sections and a few time periods. With linear models, researchers usually difference the estimating equation to remove

[1] There is a vast literature in which the spatial dependence structure is assumed known up to a finite–dimensional parameter vector. Asymptotic results for such processes can be established under the assumption of independence after an appropriate transformation. See e.g. Anselin (1988) for an early treatise. The results presented here do not presume that the spatial dependence structure is known.

[2] For instance, in oligopoly applications, the strength of dependence between two products is often determined, not only by the distance between them in product–characteristic space, but also by the proximity of competing products. More problematic than nonstationarity is the fact that the characteristics of the spatial process can depend on the number of observations. For instance, when a firm introduces a new product into a market, the nature of competition between the existing products changes; the new product is not simply a theretofore unobserved point of the SUP. Finally, both the location of economic observations and the structure of the market in which they are located can be endogenous. For instance, gas stations enter at busy intersections, manufacturers introduce products with attributes that appeal to consumers, and so forth.

the influence of time–invariant cross–sectional effects. With discrete–choice models, however, the situation is more complex. For this reason, attention is often limited to static conditional–logit models with independent errors and strictly exogenous regressors (e.g., Chamberlain 1984). Honoré and Kyriazidou (2000) generalize that model to include a lagged dependent variable but maintain the strict–exogeneity and independence assumptions, whereas Magnac (2002) considers dependent errors with arbitrary known marginal distributions but maintains the strict–exogeneity and static assumptions. We in contrast consider a dynamic discrete–choice model with endogenous regressors and arbitrary patterns of spatial and time–series dependence.

Unlike the above–mentioned papers, our fixed effect is not included in the latent–variable equation. Instead, it enters linearly in the observed–choice equation. As noted above, estimating dynamic discrete–choice models with fixed effects in the latent–variable equation is problematic and normally requires strong assumptions. Furthermore, sometimes the approach taken is nonparametric, which imposes practical limitations (e.g. data requirements, continuous regressors) in small and moderate–sized samples, even if not all of those limitations are borne out by the asymptotic distribution. In our model, like in linear panel data models,[3] the fixed effects enter linearly and can hence be removed by differencing. The interpretation of our fixed effects, however, is different. Indeed, the fixed effects in our model affect the probability of choosing a particular option directly instead of doing so indirectly via the latent variable.[4]

We apply our procedure to the estimation of flexible operating rules for mine openings and closings, which we model in a real–options context. This is a two–state optimal–switching decision problem in which a mine can be either active or inactive, and the operator must decide whether to operate the mine or to let it lie idle. We estimate a reduced–form discrete–choice equation that embodies many of the predictions of the theory of real options. The equation that we specify is similar to the one that is used in Moel and Tufano (2002), which is itself based on the theoretical model of Brennan and Schwartz (1985). In particular, we impose a Markov structure on the estimation. In other words, instead of estimating the probability that a mine is open (closed), we estimate transition probabilities (i.e., the probability of being in state k in period t, conditional on having been in state j in period $t-1$).

We use data on prices, costs, reserves, capacity, output, and technology for a panel of twenty one copper mines — projects that are both irreversible and uncertain. The panel includes all Canadian mines that operated during some portion of the period between 1980 and 1993 in which copper was the primary commodity. About two thirds of the observations pertain to periods in which the mine was active, whereas the remainder are inactive observations.

Since we model decision rules in a state–space context, the mine status at the beginning of the period is an important determinant of the current–period operating decision. This means that our estimating equations contain a temporally lagged dependent variable. Furthermore, the coefficients of some of the explanatory variables are predicted to differ in both magnitude and sign, depending on the prior state (i.e., on the lagged dependent variable). To illustrate, the theory of real options predicts that high price volatility tends to delay decisions. This means that high volatility causes the probability that a mine will be active to increase (decrease) if it was active (inactive) in the previous period. For this reason, we estimate decision rules in which the coefficients can be state dependent.

Our discrete–choice equations are not structural. Indeed, our intent is to test the predictions of more than one theory in a unified framework. Therefore, rather than imposing restrictions that are

[3] See, e.g., Arellano and Honoré (2001) for a survey of linear panel data models

[4] The interpretation of the fixed effects varies with the exogeneity assumptions.

implied by theories that might not be valid, we attempt to distinguish among theories by examining whether the data are consistent with their predictions.

To anticipate, we find little support for the real–options model. In particular, the signs of coefficients (e.g., the effects of volatility) do not vary with the prior state. Instead, a more conventional mean/variance utility model receives more support. We also find that, although our spatial state–dependent models have greater predictive power, they are associated with reductions in significance *vis–à–vis* an ordinary probit.

The organization of the paper is as follows. The next section deals with estimation. In particular, it presents our nonlinear dynamic panel–data model, describes our central limit theorem, and discusses our continuous–updating GMM estimation technique. Section 3 deals with the application. That section briefly discusses the testable predictions that can be derived from the theory of real options, and it describes the industry and the data. Section 4 presents estimates of static and dynamic ordinary–probit and spatial discrete–choice models, and section 5 concludes. Proofs are contained in the appendix.

2 Econometric Methodology

2.1 Our Panel Data Model

Our model is a dynamic space–time discrete–choice equation with fixed effects, i.e.

$$y_{it} = I(x'_{it1}\theta_0 - \epsilon_{it1} \geq 0)y_{i,t-1} + I(x'_{it0}\theta_0 - \epsilon_{it0} \geq 0)(1 - y_{i,t-1}) + \eta_i + u^*_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \tag{1}$$

where y_{it} is the binary choice of firm i at time t, η_i is a fixed effect, the ϵ_{itj}'s and u^*_{it}'s are errors, θ_0 is an unknown vector of regression coefficients, x_{it1} and x_{it0} are regressor vectors and I is the indicator function.

Model (1) allows for various regressor configurations. If $x_{it1} = x_{it0}$, then the model reduces to a static one. If $x_{it1} = [x'_{it}, 0']'$ and $x_{it0} = [0', x'_{it}]$ then the regressors in both components of (1) are the same but the regression coefficients are allowed to be different. Finally, any combination of the two extremes is possible. We use x_{it} to denote all regressors that are in at least one of x_{it1}, x_{it0}.

We assume that i) the ϵ_{itj}'s have standard normal distributions; ii) ϵ_{itj} is independent of $y_{i,t-1}$;[5] iii) the ϵ_{itj}'s are independent of current and past x_{itj}'s;[6] iv) a vector of instruments z_{it} exists that are independent of the ϵ_{itj}'s and for which $E(u^*_{it}|z_{it}) = E(u^*_{i,t-1}|z_{it}) = 0$ a.s.[7] Typically z_{it} would consist of regressors lagged at least one period.

Now,

$$E(y_{it}|z_{it}) = E\Big(I(x'_{it1}\theta_0 - \epsilon_{it1} \geq 0)y_{i,t-1}|z_{it}\Big) + E\Big(I(x'_{it0}\theta_0 - \epsilon_{it0} \geq 0)(1 - y_{i,t-1})|z_{it}\Big) + E(\eta_i|z_{it}). \tag{2}$$

[5]This can be done without loss of generality since ϵ_{itj} can be written as

$$\epsilon_{itj} = \epsilon_{itj;1}y_{i,t-1} + \epsilon_{itj;0}(1 - y_{i,t-1}),$$

with $\epsilon_{itj;s}$ independent of $y_{i,t-1}$ for all i, j, t, s. Then ϵ_{itj} in (1) can be simply replaced with $\epsilon_{itj;j}$, which is independent of $y_{i,t-1}$.

[6]Please note that regressors are random here, not deterministic. There is hence no implicit strict exogeneity assumption and condition iii) is not implied by condition i) since condition i) relates to the unconditional error distribution, not the error distribution conditional on the regressors.

[7]This means that endogeneity is due to the fact that regressors can be correlated with u^*, not with $\epsilon_j, j = 0, 1$.

But

$$E\Big(I(x'_{it1}\theta_0 - \epsilon_{it1} \geq 0)y_{i,t-1}|z_{it}\Big) = E\Big(E\big(I(x'_{it1}\theta_0 - \epsilon_{it1} \geq 0)y_{i,t-1}|z_{it}, y_{i,t-1}, x_{it}\big)|z_{it}\Big)$$
$$= E\Big(\Phi(x'_{it1}\theta_0)y_{i,t-1}|z_{it}\Big) \text{ a.s..}$$

Repeat the same steps for the second right hand side term in (2) to obtain

$$E(y_{it}|z_{it}) = E\Big(\Phi(x'_{it1}\theta_0)y_{i,t-1}|z_{it}\Big) + E\Big(\Phi(x'_{it0}\theta_0)(1-y_{i,t-1})|z_{it}\Big) + E(\eta_i|z_{it}) \text{ a.s.}, \tag{3}$$

where Φ is the standard normal distribution function. Take first differences to obtain

$$E\Big(y_{it}-y_{i,t-1}-\Phi(x'_{it1}\theta_0)y_{i,t-1}-\Phi(x'_{it0}\theta_0)(1-y_{i,t-1})+\Phi(x'_{i,t-1,1}\theta_0)y_{i,t-2}+\Phi(x'_{i,t-1,0}\theta_0)(1-y_{i,t-2})|z_{it}\Big) = 0 \text{ a.s..}$$

Like in linear panel–data models the nature of the dependence between fixed effects and other model variables is irrelevant. However, unlike in linear panel data models, time–invariant regressors are not differenced out with the fixed effects. Let

$$g_{it}(\theta) = z_{it}\Big(y_{it}-y_{i,t-1}-\Phi(x'_{it1}\theta)y_{i,t-1}-\Phi(x'_{it0}\theta)(1-y_{i,t-1})+\Phi(x'_{i,t-1,1}\theta)y_{i,t-2}+\Phi(x'_{i,t-1,0}\theta)(1-y_{i,t-2})\Big). \tag{4}$$

Then

$$\forall i,t : Eg_{it}(\theta_0) = 0. \tag{5}$$

We thus have the main prerequisite for application of a GMM style procedure: a set of moment conditions. Although not explicit in the notation above, the g_{it}'s will be allowed to vary with N, T, and can also vary across i, t provided that (5) is satisfied.

We now state our generic theoretical results. In the technical sections that follow n is either N or NT, depending on whether T is fixed or increases.

2.2 A Suitable CLT

In Pinkse, Shen, and Slade (2005) (PSS), we develop a new CLT that is designed to address shortcomings in previously available CLT's. We now summarize the assumptions and results of that paper, which are used to establish the properties of the CUE in this paper.

CLT's are in essence results about sums of zero mean random variables. Because the statistical properties — including the strength and nature of dependence between observations — should be allowed to vary with the sample size, we index them by the sample size (i.e. our observations are $\xi_{n1}, \ldots, \xi_{nn}$ and their sum is S_n).

The idea in PSS, which is based on an idea by Bernstein (1927), is to divide the observations into nonoverlapping groups $\mathcal{G}_{n1}, \ldots, \mathcal{G}_{nJ}$, $1 \leq J < \infty$, which are divided up into mutually exclusive subgroups $\mathcal{G}_{nj1}, \ldots, \mathcal{G}_{njm_{nj}}$, $j = 1, \ldots, J$. Group membership of each observation can vary with the sample size n and so can the number of subgroups m_{nj} in group $j = 1, \ldots, J$. Partial sums over elements in groups and subgroups are denoted by S_{nj} and S_{njt}, $j = 1, \ldots, J$ and $t = 1, \ldots, m_{nj}$, respectively. Thus,

$$S_n = \sum_{j=1}^{J} S_{nj} = \sum_{j=1}^{J}\sum_{t=1}^{m_{nj}} S_{njt} = \sum_{i=1}^{n} \xi_{ni}.$$

The only role of groups 2 through J is to reduce the strength of the dependence across subgroups in group 1.

To illustrate, suppose that there are a number of gasoline stations in a city. The idea is to partition stations into 'sets' that compete intensely with one another, e.g. ones located nearby, ones located along the same thoroughfare, or ones offering similar additional services (see e.g. Pinkse and Slade, 1998). No matter how one chooses the sets, however, there will often be stations at the 'boundary' of one set that face strong competition from a station in another set. However, if set 2 is located between sets 1 and 3, then competition between stations in set 1 and those in set 3 is likely to be small.

As the city grows, the number of stations will also grow, and new stations will appear both at the periphery and, due to increased population density, also in established areas. Indeed, wherever entry is deemed profitable, be it because competition is weak or because the market is large, expansion will occur. Furthermore, other stations will shut down because, for example, the land is more valuable in alternate uses or because they have become unprofitable. This means that, as the city grows, the choice of sets will change.

The idea, then, is that each of the sets is a subgroup and that subgroups are allocated to groups in such a way that dependence between observations in different subgroups of the same group is small. With the example, sets 1 and 3 could be subgroups of the same group, whereas set 2 would be in a different group. As the city grows, it is possible to allocate ever more stations to each subgroup. Moreover, as this process continues, the level of competition between stations in different subgroups of the same group will dissipate due to increased competition from stations of another group that are located between them. In the limit, dependence will disappear altogether and we will be back in a familiar situation of independent random variates.

In PSS, we combine this grouping idea with a weak–dependence assumption that is due to Doukhan and Louhichi (1999). That assumption is weaker than strong mixing (Rosenblatt, 1956) and easier to work with than near–epoch dependence (Ibragimov, 1962). We now state the assumptions and results of PSS without further elaboration.[8]

Definition 1 *Let $\mathcal{F}$ be a collection of functions $\{f : \forall t \in \mathbb{R} : f(t) = t$ or $\exists u \in \mathbb{R} : \forall t : f(t) = e^{\iota u t}\}$, where ι is the imaginary number.*

Assumption A *For any $j = 1, \ldots, J$, let $\mathcal{G}_n^*, \mathcal{G}_n^{**} \subset \mathcal{G}_{nj}$ be any sets for which*

$$\forall t = 1, \ldots, m_{nj} : \mathcal{G}_{njt} \cap \mathcal{G}_n^* \neq \emptyset \Rightarrow \mathcal{G}_{njt} \cap \mathcal{G}_n^{**} = \emptyset.$$

Then for any function $f \in \mathcal{F}$,

$$\left\| Cov\left(f\Big(\sum_{s\in\mathcal{G}_n^*} \xi_{ns}\Big), f\Big(\sum_{s\in\mathcal{G}_n^{**}} \xi_{ns}\Big)\right)\right\| \leq \sqrt{Vf\Big(\sum_{s\in\mathcal{G}_n^*} \xi_{ns}\Big)}\sqrt{Vf\Big(\sum_{s\in\mathcal{G}_n^{**}} \xi_{ns}\Big)}\alpha_{nj}, \tag{6}$$

for some 'mixing' numbers α_{nj} with

$$\lim_{n\to\infty} \sum_{j=1}^{J} m_{nj}^2 \alpha_{nj} = 0. \tag{7}$$

[8] A detailed description can be found in PSS, section 2.3.

Assumption B

$$\lim_{n\to\infty} \max_{t\le m_{nj}} \sigma_{njt}/\varsigma_{nj} = 0,\ j = 1,\dots,J, \qquad \lim_{n\to\infty} \varsigma_{nj}/\varsigma_{n1} = 0,\ j = 2,\dots,J. \tag{8}$$

Assumption C *For some sequence* $\{h_n\}$ *for which* $\lim_{n\to\infty} h_n = 0$ *and for all* $j = 1,\dots,J$,

$$\lim_{n\to\infty} \max_{t\le m_{nj}} E\left(\frac{S_{njt}^2}{\sigma_{njt}^2} I\Big(\Big|\frac{S_{njt}}{\varsigma_{nj}}\Big| > h_n\Big)\right) = 0. \tag{9}$$

A sufficient condition for assumption C is that for some $p > 1$,

$$E|S_{njt}|^{2p} = o\big(\sigma_{njt}^2 \varsigma_{nj}^{2p-2}\big), \quad j = 1,\dots,J;\ t = 1,\dots,m_{nj}.^{9} \tag{10}$$

We can now state our theorems.

Theorem 1 *If assumptions A–C hold, then*

$$\frac{S_n}{\sigma_n} \xrightarrow{\mathcal{D}} N(0,1). \tag{11}$$

A vector–valued version of 1 is also available. Let $\vec{\xi}_{ni}$ be vector–valued and

$$\vec{S}_n = \sum_{i=1}^{n} \vec{\xi}_{ni}.$$

Let furthermore $\Sigma_n = V\vec{S}_n$.

Theorem 2 *If for any vector* v *with* $||v|| = 1$, *assumptions A–C are satisfied for* $\xi_{ni} = v'\Sigma_n^{-1/2}\vec{\xi}_{ni}$, *then*

$$\Sigma_n^{-1/2}\vec{S}_n \xrightarrow{\mathcal{D}} N(0, I). \tag{12}$$

2.3 Continuous Updating — One–step GMM

The continuous updating estimator is similar to the regular two–step GMM estimator albeit that the weight matrix is parametrized immediately. Our moment condition is

$$\forall i, n : Eg_{ni}(\theta_0) = 0,$$

where $\theta_0 \in \Theta \subset \mathbb{R}^d$ is the vector of parameters of interest, and g_{ni} is some vector–valued function. The CUE is

$$\hat{\theta} = \operatorname{argmin}_{\theta\in\Theta} \hat{\Omega}_n(\theta).$$

where the CUE objective function $\hat{\Omega}_n$ has the form

$$\hat{\Omega}_n(\theta) = \bar{g}_n'(\theta)\hat{W}_n(\theta)\bar{g}_n(\theta),$$

[9] By the Hölder and Markov inequalities, $E\big(S_{njt}^2 I(|S_{njt}| > h_n\varsigma_{nj})\big) \le (E|S_{njt}|^{2p})^{1/p}\big(P(|S_{njt}| > h_n\varsigma_{nj})\big)^{1-1/p} \le E|S_{njt}|^{2p} h_n^{2-2p}\varsigma_{nj}^{2-2p}$.

with

$$\bar{g}_n(\theta) = n^{-1} \sum_{i=1}^{n} g_{ni}(\theta), \tag{13}$$

$$\hat{W}_n(\theta) = \Psi_n \hat{V}_n^{-1}(\theta), \tag{14}$$

where $\{\Psi_n\}$ is a sequence of numbers to be defined below and

$$\hat{V}_n(\theta) = n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \big(g_{ni}(\theta) - \bar{g}_n(\theta)\big)\big(g_{nj}(\theta) - \bar{g}_n(\theta)\big)' \tag{15}$$

is such that $\hat{V}_n(\theta_0)$ is an estimator of the asymptotic variance of $\sqrt{n}\bar{g}_n(\theta_0)$. Since the Ψ_n's in (14) are scalars their inclusion does not affect the estimates, but they facilitate the proofs. The numbers λ_{nij} in (15) are weights; if observations are known to be independent only the λ_{nii}'s need to be nonzero; their choice is discussed below. We follow Hansen, Heaton and Yaron (1996) in having the $\bar{g}_n$'s in (15) but our results will also go through if they are omitted. Their main purpose is practical, i.e. to avoid having $\hat{V}_n$ very large (and $\hat{W}_n$ very small) when θ is far from θ_0.

In large samples, the objective function $\hat{\Omega}_n$ is close to Ω_n defined by

$$\Omega_n(\theta) = g_n'(\theta) W_n(\theta) g_n(\theta),$$

where $g_n(\theta) = E\bar{g}_n(\theta)$, $W_n(\theta) = \Psi_n V_n^{-1}(\theta)$ and

$$V_n(\theta) = n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij}\big(g_{ni}(\theta) - g_n(\theta)\big)\big(g_{nj}(\theta) - g_n(\theta)\big)'\Big).$$

So provided that $g_n(\theta) = 0$ if and only if $\theta = \theta_0$ and that W_n is a positive definite matrix, $\Omega_n(\theta) = 0 \Leftrightarrow \theta = \theta_0$.

Consistency

A number of conditions are necessary for consistency of our CUE, which we now explain. Recall that we do not assume stationarity (i.e. we neither assume that observations are located on a regular grid nor that dependence is equally strong between all pairs of observations that are equally far apart) and that we allow for the dependence structure to change with the sample size. For these reasons, our conditions are more difficult to express, (i.e. they are more technical) than most. However, at the end of the subsection, we include a discussion of the implications of our assumptions in the context of a simpler spatial setting as well as for our dynamic discrete–choice model.

We now state the conditions necessary for consistency, followed by a discussion of each.

Assumption D *θ_0 is an interior point of Θ, which is convex and compact.*

The compactness portion of assumption D is standard. Convexity is somewhat unusual, but is reasonable in most applications.

Let Λ_n be the $n \times n$ matrix with i, j element λ_{nij}

Assumption E *For some deterministic sequence* $\{\chi_n\}$ *with* $\chi_n = O(1)$,

$$\operatorname{ess\,sup} \max_{i\leq n} E\big(\max_{\theta\in\Theta} ||g_{ni}(\theta)||^4|\Lambda_n\big) \leq \chi_n,^{10} \tag{16}$$

$$\max_{i\leq n} E \max_{\theta\in\Theta} \Big|\Big|\frac{\partial g_{ni}}{\partial \theta_s}(\theta)\Big|\Big|^2 \leq \chi_n, \quad s = 1,\ldots,d. \tag{17}$$

We condition on Λ_n in (16) since the weights λ_{nij} can be random.

We need to ensure that θ_0 is the unique solution to $g_n(\theta) = 0$ for any sufficiently large n, which is accomplished by assumption F.

Assumption F *For some continuous function* $g^* : \Theta \to \mathbb{R}$,

$$\exists n^* : \forall n > n^*, \theta \in \Theta : ||g_n(\theta)|| \geq g^*(\theta) \text{ and } g^*(\theta) = 0 \Leftrightarrow \theta = \theta_0.$$

We will also put some restrictions on the strength of the dependence. Let $\{\alpha_{nij}\}$ be some numbers that satisfy

$$\begin{cases} \Big|\mathrm{Cov}\big(g_{nit}(\theta), g_{njs}(\theta)\big)\Big| & \leq \alpha_{nij}\sqrt{Vg_{nit}(\theta)}\sqrt{Vg_{njs}(\theta)}, \\ \Big|\mathrm{Cov}\big(\gamma_{nij}(\theta), \gamma_{ni^*j^*}(\theta)\big)\Big| & \leq \big(\alpha_{nii^*} + \alpha_{nij^*} + \alpha_{nji^*} + \alpha_{njj^*}\big)\sqrt{V\gamma_{nij}(\theta)}\sqrt{V\gamma_{ni^*j^*}(\theta)}, \end{cases} \tag{18}$$

for all $i, j, i^*, j^*; \theta \in \Theta$, where γ_{nij} is one of $\lambda_{nij}, \lambda_{nij}g_{nit}, \lambda_{nij}g_{nit}g_{njs}$, for any $t, s = 1,\ldots,d_g$, where d_g is the dimension of the g_{ni}–vector.

Without conditions on the α–coefficients, (18) is not restrictive. Restrictions will be put on by means of the rate of increase of the sequence $\{A_n\}$ defined by

$$A_n = \max_{i\leq n} \sum_{j=1}^{n} \alpha_{nij} \tag{19}$$

In most applications assuming that $\limsup_{n\to\infty} A_n$ is finite is reasonable, but for the consistency proof we allow A_n to grow with n. The conditions imposed on A_n are stated in assumption I.

We also need to impose some conditions to ensure that $\hat{V}_n(\theta) - V_n(\theta)$ vanishes and that $V_n(\theta)$ is always invertible. To achieve this, we have to make assumptions on the choice of weights λ_{nij} and on the variability in the g_{ni}'s.

Assumption G *For some deterministic sequence* $\{\zeta_n\}$ *with* $\zeta_n = O(1)$,

$$\min_{\theta\in\Theta} \mathcal{E}_{\min}\Big(n^{-1}\sum_{i=1}^{n} Vg_{ni}(\theta)\Big) \geq \zeta_n^{-1}, \quad \max_{\theta\in\Theta} \mathcal{E}_{\max}\Big(n^{-1}\sum_{i=1}^{n} Vg_{ni}(\theta)\Big) \leq \zeta_n,$$

where $\mathcal{E}_{\min}, \mathcal{E}_{\max}$ *denote the minimum and maximum eigenvalue of a matrix.*

The nonnegative weights λ_{nij} should ideally be chosen such that λ_{nij} is equal or close to one when observation i's location is close to j's. We suggest a choice for these weights at the end of this section. Without loss of generality we will assume the matrix Λ_n with elements λ_{nij} to be symmetric.

[10] The essential supremum of a random variable ω is $\min\{t : P(|\omega| > t) = 0\}$ (see e.g. (8.1) in Davidson (1994)).

Assumption H *Deterministic sequences of nonnegative numbers* $\{\rho_n\}, \{\Psi_n\}$ *exist such that*

$$\Psi_n = \max\left\{\max_{i\leq n}\sum_{j=1}^{n}\sqrt{E\lambda_{nij}^2}, \text{ess sup}\big(\mathcal{E}_{\max}(\Lambda_n)\big)\right\}, \tag{20}$$

$$\rho_n = \text{ess sup}\,\frac{1}{\mathcal{E}_{\min}(\Lambda_n)}, \tag{21}$$

and $\rho_n = O(1), \Psi_n^{-1} = O(1)$.

If Λ_n is singular, then V_n may also not be invertible. Note that the assumptions on Λ_n required for consistency are weak; the shape of Λ_n is not important.

Assumption I A_n, Ψ_n *are such that*

$$\lim_{n\to\infty} n^{-1}A_n\Psi_n^{2(2+d)}\mathcal{L}_n^{2+d} = 0, \tag{22}$$

for some $\mathcal{L}_n$ *with* $\lim_{n\to\infty}\mathcal{L}_n = \infty$.

Even if $\lim_{n\to\infty} A_n < \infty$, Ψ_n can increase only slowly with n, certainly more slowly than is necessary for the 'optimal' rate of increase of the cutoff parameter of the Newey–West estimator. It is possible to weaken the rate of increase limitations on Ψ_n by strengthening other conditions. The current limitation arises from an apparently new generic uniform convergence result, see lemma 1 in the appendix.

Theorem 3 *If assumptions D–I hold then* $\hat{\theta} \xrightarrow{\mathcal{P}} \theta_0$.

Note that theorem 3 requires no weak dependence conditions over and above that implied in the requirements imposed on A_n (defined in (19)) in assumption I. So no groups and such need to be chosen, the only restriction is on the covariances.

Asymptotic Normality

The conditions required for asymptotic normality are stronger than those needed for consistency. The objective is to show that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N\big(0, (T_0'V_0^{-1}T_0)^{-1}\big), \tag{23}$$

where

$$T_0 = \lim_{n\to\infty}\frac{\partial g_n}{\partial\theta'}(\theta_0), \quad V_0 = \lim_{n\to\infty} V_n^*(\theta_0), \text{ with } V_n^*(\theta_0) = n^{-1}\sum_{i,j=1}^{n} E\Big(g_{ni}(\theta_0)g_{nj}'(\theta_0)\Big). \tag{24}$$

We need to make assumptions to ensure that the limits in (24) exist, that $\frac{\partial\bar{g}_n}{\partial\theta'}(\hat{\theta}^*)$ and $\hat{V}_n(\hat{\theta}^*)$ converge to T_0 and V_0 whenever $\hat{\theta}^*$ is a consistent estimator of θ_0, that the asymptotic variance matrix in (23) is well–defined, and that the conditions of theorem 2 are satisfied.

Assumption J *The matrices* T_0, V_0 *defined in (24) are finite and have maximum (column) rank.*

Assumption K $\{g_{ni}(\theta_0)\}$ *satisfies the assumptions of theorem 2.*

Now let

$$\left|\mathrm{Cov}\Big(\frac{\partial g_{ni;t^*}}{\partial \theta_t}(\theta), \frac{\partial g_{nj;s^*}}{\partial \theta_s}(\theta)\Big)\right| \leq \alpha_{nij}\sqrt{V\frac{\partial g_{ni;t^*}}{\partial \theta_t}(\theta)}\sqrt{V\frac{\partial g_{nj;s^*}}{\partial \theta_s}(\theta)}, \tag{25}$$

for all i, j, t, s, t^*, s^* and all $\theta \in \Theta$. A_n remains defined as in (19), but note that with the new requirement on the α–coefficients, the requirements imposed on A_n in assumption I is now somewhat stronger. Note further that the strength of the dependence is now also controlled indirectly through assumption K. The assumptions for theorem 2, imposed by assumption K, only apply to (sums of) the g_{ni}'s evaluated at θ_0, but they apply both to the sums of g_{ni}'s themselves and to complex exponentials thereof. The conditions imposed on the rate of increase of A_n control the covariance sums of the g_{ni}'s and $\partial g_{ni}/\partial\theta'$, uniformly in θ. Both sets of assumptions are weak.

The next assumption imposes some additional constraints on the choice of weights λ_{nij}.

Assumption L *The weights* $\{\lambda_{nij}\}$ *are such that for* $t, s = 1, \ldots, d_g$.

$$\lim_{n\to\infty} n^{-1}\sum_{i,j=1}^{n} E\big|(1-\lambda_{nij})g_{ni;t}(\theta_0)g_{nj;s}(\theta_0)\big| = 0.$$

Assumption L suggests that the practitioner must have some information on the dependence. The information need not be detailed but she must have some idea for which (i, j)–pairs g_{ni} and g_{nj} are highly correlated. Note that assumption L only applies to the g_{ni}'s at θ_0, not at arbitrary values.

Assumption M

$$\{\frac{\partial \bar{g}_n}{\partial \theta'}\} \text{ is uniformly stochastically equicontinuous on } \Theta.$$

Uniform stochastic equicontinuity is a technical assumption which is implied by all partial second derivatives of g_{ni} being bounded in probability, uniformly in $\theta \in \Theta$.

Theorem 4 *If* $V_0 > 0$ *and assumptions D–M are satisfied, then (23) holds.*

Bias Correction

Unlike the GMM estimator, the CUE is a generalized empirical likelihood (GEL) estimator. Consequently, a second order bias correction is possible, where such a correction is not feasible for the two–step estimator. The bias correction does not affect the asymptotic distribution of the CUE, so the CUE generally has the same asymptotic distribution as the two–step GMM estimator. But a bias correction can improve performance in moderate size samples.

To see how a second order bias correction can help, note the following. The difference $\hat{\theta} - \theta_0$ can be expanded into terms converging at rates $n^{-1/2}$, n^{-1}, $n^{-3/2}$, etcetera, i.e. the first, second, third, ... terms in an asymptotic expansion. The first order term is responsible for the asymptotic distribution; all other terms are of no consequence in the limit. However, the other expansion terms can make a difference in samples of finite size. By reducing the bias in the second order term, the precision of the estimator is improved.

We do not provide such a bias correction ourselves. However, in work motivated by our current paper, Iglesias and Phillips (2005) provide this bias correction for models with the nonstationarity problem we address here, not only for the CUE but also for the empirical likelihood estimator.

Conditions in a Simple Spatial Model

We now offer a a further explanation of our assumptions in a nonlinear regression model with the simplest possible spatial dependence structure: one with stationary data that are equally spaced on a line and where the locations equal the observation indices. An example of such a process can be found in Whittle (1954); Cliff and Ord (1973) describe similar models. This model does not do justice to the generality of our results but it does help clarify the assumptions. Although the assumptions on the dependence structure are somewhat milder here than in e.g. Conley (1999), the moment conditions implied by our assumptions are probably somewhat stronger than necessary, but hardly unreasonable, in this simple case. Since the dependence structure is independent of the sample size in this model, we drop the dependence on n in our notation here.

Consider the nonlinear regression model

$$y_i = h(x_i, \theta_0) + u_i = h_i(\theta_0) + u_i, \quad i = 1, \ldots, n,$$

where the y's, x's and u's used here are unrelated to those used in our dynamic spatial discrete choice model. Suppose that the dimensions of x_i and θ_0 are both d. If the regressors x_i are used as instruments, then g would be given by

$$g_i(\theta) = x_i\big(y_i - h(x_i, \theta)\big), \quad i = 1, \ldots, n.$$

The locations are nonrandom here so the λ_{ij}'s drop out of any expectation. For instance, assumption E simplifies to

$$E \max_{\theta \in \Theta} \big|\big| x_1\big(y_1 - h_1(\theta)\big) \big|\big|^4 < \infty, \quad E \max_{\theta \in \Theta} \left|\left| x_1 \frac{\partial h_1}{\partial \theta_s}(\theta) \right|\right|^2 < \infty,$$

for all $s = 1, \ldots, d$. Further, assumption G becomes $0 < Vg_1(\theta) < \infty$ and assumption M would be implied by (see e.g. Davidson (1994), theorem 21.10),

$$E\left(\max_{\theta \in \Theta} \left|\left| \frac{\partial^2 g_{1t}}{\partial \theta \partial \theta'}(\theta) \right|\right| \right) < \infty, \quad t = 1, \ldots, d.$$

Assumption F is implied by the usual identification condition on θ_0 in the nonlinear regression model specified and assumption J requires that $E(u_1^2 x_1 x_1')$ > 0 and $E\big(x_1 \partial h_1 / \partial \theta'(\theta_0)\big)$ is invertible.

If the x_i's were moreover nonrandom then (18) would simplify to

$$|\mathrm{Corr}(u_i, u_j)| \leq \alpha_{ij}, \quad |\mathrm{Corr}(u_i u_j, u_{i^*} u_{j^*})| \leq \alpha_{ii^*} + \alpha_{ij^*} + \alpha_{ji^*} + \alpha_{jj^*},$$

for all i, j, i^*, j^*. We find it difficult, however, to think of an obvious application for fixed regressors in economics. With random regressors, (18) applies to a combination of errors and regressors, i.e. (18) would be equivalent to

$$\Big|\mathrm{Corr}\big(g_{it}(\theta), g_{js}(\theta)\big)\Big| \leq \alpha_{ij}, \quad \Big|\mathrm{Corr}\big(g_{it}(\theta) g_{js}(\theta), g_{i^*t^*}(\theta) g_{j^*s^*}(\theta)\big)\Big| \leq \alpha_{ii^*} + \alpha_{ij^*} + \alpha_{ji^*} + \alpha_{jj^*},$$

for all i, j, i^*, j^* and all $\theta \in \Theta$. Condition (25) only adds the requirement that

$$\left|\mathrm{Cov}\Big(x_{it^*} \frac{\partial h_i}{\partial \theta_t}(\theta), x_{js^*} \frac{\partial h_j}{\partial \theta_s}(\theta)\Big)\right| \leq \alpha_{ij} \sqrt{V\Big(x_{it^*} \frac{\partial h_i}{\partial \theta_t}(\theta)\Big)} \sqrt{V\Big(x_{js^*} \frac{\partial h_j}{\partial \theta_s}(\theta)\Big)},$$

for all i, j, t, s, t^*, s^*.

By stationarity, $\alpha_{ij} = \alpha_{i^*j^*}$ for any i, j, i^*, j^* for which $|i-j| = |i^*-j^*|$. Thus, we could replace the notation α_{ij} by α_{i-j}. (18) and (25) are implied by a *strong mixing* assumption with mixing numbers α_t. If the (strong) mixing numbers are moreover summable then $A_n = O(1)$.

Now, with respect to the choice of λ–weights, (20) now defines Ψ_n as the maximum row sum of the Λ_n matrix and requires its smallest eigenvalue to be bounded away from zero in the limit. Given the dependence structure imposed it would be natural to select $\lambda_{ij} = \lambda_{i-j}$, possibly as $\lambda_t = \big(1 - |t|/(2\kappa_n+1)\big) I(|t| \leq \kappa_n)$ where κ_n increases as a small fractional power of n. Clearly then $\Psi_n \leq \kappa_n$. For this choice of λ's, assumption L is implied by

$$n^{-1} \sum_{i,j=1}^{n} (1 - \lambda_{i-j}) \alpha_{i-j} = o(1),$$

which after some tedious algebra can in turn be shown to be implied by

$$\sum_{i=2\kappa_n}^{\infty} \alpha_i + \kappa_n^{-1} \sum_{j=0}^{2\kappa_n} j \alpha_j = o(1),$$

which can be seen to hold if $A_n = O(1)$.

Assumption I then requires that $n^{-1} \kappa_n^{2(2+d)} \mathcal{L}_n^{2+\delta} = o(1)$, which can be made to hold by letting κ_n increase suitably slowly.

Finally we discuss assumption K, which requires assumptions A–C. We can use the grouping scheme of figure 1. We choose $J = 2$ groups, each with $m_{n1} = m_{n2}$ subgroups. The subgroups of group j have ℓ_{nj}, $j = 1, 2$, observations each. So observations $1, \ldots, \ell_{n1}$ belong to subgroup 1 of group 1, the next ℓ_{n2} belong to subgroup 1 of group 2, the following ℓ_{n1} to subgroup 2 of group 1, and so forth.

We assume here that the mixing weights decline at a rate of $\alpha_t \sim t^{-w}$ for some $w > 0$. First, noting that $\ell_{n2} = o(\ell_{n1})$ by assumption B, (7) requires that $m_{n2}^2 \alpha_{\ell_{n2}} = o(1)$, i.e. $m_{n2}^2 \ell_{n2}^{-w} = o(1)$.

If w is large then we can afford to choose many subgroups with few observations each. For instance, if $w > 4$, then we can have $m_{n1} = m_{n2} = n^{2/3}$, $\ell_{n2} = n^{(1+4/w)/6}$ and $\ell_{n1} = n^{1/3} - \ell_{n2}$.[11] Because S_{njt} is a summation over ℓ_{nj} observations, here $\sigma_{njt}^2, \varsigma_{nj}^2$ will be of order no less than ℓ_{nj} and $m_{nj}\ell_{nj}$, respectively.[12] Although $E|S_{njt}|^{2p}$, $p > 2$, will often be $O(\ell_{nj}^p)$, it is certainly $O(\ell_{nj}^{2p})$ and since $\ell_{nj}^{2p}/(\ell_{nj}^p m_{nj}^{p-1}) = \ell_{nj}^p m_{nj}^{1-p} = O(n^{p/3} \cdot n^{2(1-p)/3}) = O(n^{(2-p)/3}) = o(1)$, (10) is satisfied. Similarly, if the number of existing moments is large, then we choose fewer subgroups with more elements each and w only needs to be slightly larger than 2. If the weak dependence conditions are strengthened and weak dependence within each subgroup is exploited then values of $w < 1$ are achievable; we do not show this here.

Conditions in the Dynamic Discrete Choice Model

It is instructive to discuss the meaning of our conditions in the context of our dynamic discrete choice panel–data model. Several assumptions, including those relating to weak dependence and the choice of weights, are largely independent of the particular case of our dynamic discrete choice model, and assumption F is satisfied if a sufficiently large number of instruments is available. We

[11] We ingore the possibility that these numbers may not be integers here.

[12] as they would be in the i.i.d. case.

will therefore focus our attention on assumptions E, G, and M. Recall that in our dynamic discrete choice model the dependence of the variables on n was not evident in the notation; here we continue with that notation and use y_{it} instead of y_{nit}, etcetera. Like in the example above in this discussion we again assume locations are nonrandom and hence so are the λ's.

First consider assumption E, and note that $\Phi, \phi = \Phi', \phi'$ are all bounded, i.e. for some $C_\phi < \infty$

$$\max_{t\in\mathbb{R}}\big(\Phi(t) + \phi(t) + |\phi'(t)|\big) \le C_\phi.$$

Now, $|y_{it}| \le 1$, and hence

$$\begin{aligned}\max_{\theta\in\Theta} ||g_{it}(\theta)|| \le ||z_{it}|| \times \max_{\theta\in\Theta}\Big(y_{it} + \big(1 + |\Phi(x'_{it1}\theta)| + |\Phi(x'_{it0}\theta)|\big)y_{i,t-1} + \big(|\Phi(x'_{i,t-1,1}\theta)| + |\Phi(x'_{i,t-1,0}\theta)|\big)y_{i,t-2} \\ + \big(|\Phi(x'_{it0}\theta)| + |\Phi(x'_{i,t-1,0}\theta)|\big)\Big) \\ \le ||z_{it}|| \times (2 + 6C_\phi).\end{aligned}$$

Hence condition (16) of assumption E is satisfied when $E||z_{it}||^4$ is bounded uniformly in i, t, n. The uniformity condition is only needed because the expectations are potentially different and can change with the number of observations.

For condition (17) the derivation is similar albeit that the $\Phi(x'\theta)$'s in the last–displayed equation are replaced with $x_s\phi(x'\theta)$'s. Thus,

$$\max_{\theta\in\Theta}\Big|\Big|\frac{\partial g_{it}}{\partial\theta_s}(\theta)\Big|\Big| \le ||z_{it}|| \times \Big(2 + C_\phi\big(|x_{it1s}| + 2|x_{it0s}| + |x_{i,t-1,1s}| + 2|x_{i,t-1,0s}|\big)\Big)$$

and (17) holds if $E||z_{it}||^4$, $E||x_{it1}||^4$, and $E||x_{it0}||^4$ are bounded uniformly in i, t, n.

The second part of assumption G is implied by (16) and the first half holds for stationary data unless the instruments are linearly dependent, as noted earlier.

Finally, using the comment following assumption M, assumption M follows from the fact that

$$\begin{aligned}&\max_{\theta\in\Theta}\Big|\Big|\frac{\partial^2 g_{it}}{\partial\theta_s\partial\theta_{s^*}}(\theta)\Big|\Big| \le \\ &||z_{it}|| \times \Big(2 + C_\phi\big(|x_{it1s}| + 2|x_{it0s}| + |x_{i,t-1,1s}| + 2|x_{i,t-1,0s}|\big)\big(|x_{it1s^*}| + 2|x_{it0s^*}| + |x_{i,t-1,1s^*}| + 2|x_{i,t-1,0s^*}|\big)\Big),\end{aligned}$$

provided that $E||z_{it}||^6$, $E||x_{it1}||^6$ and $E||x_{it0}||^6$ are bounded uniformly in i, t, n.

2.4 Choice of Weights

We propose a scheme for choosing the λ–weights, which is in the spirit of Newey and West (1987). But the Newey–West weights depend on the stationarity assumption as do other weighting schemes for spatial estimators proposed by Conley (1999) and Kelejian and Prucha (2004). Our weights are

$$\lambda_{nij} = \frac{\sum_{t=1}^n \tau_{nti}\tau_{ntj}}{\Big(\sum_{t=1}^n \tau_{nti}^2 \sum_{t=1}^n \tau_{ntj}^2\Big)^{1/2}}, \tag{26}$$

where the τ_{nti} are numbers that are large when observation i is near t. The matrix Λ_n with (i, j) element λ_{nij} as defined in (26) is necessarily positive semidefinite since it is $\Lambda_n = \sum_{t=1}^n \tilde\tau_{nt}\tilde\tau'_{nt}$, where

$\tilde{\tau}_{nt}$ is a vector with i–th element $\tau_{nti}/\sqrt{\sum_{t=1}^{n}\tau_{nti}^2}$. If Λ_n is positive semidefinite, then so is $\hat{V}_n$,[13] which is necessary for the one–step GMM procedure to work.

Expression (26) is generic. In a stationary time series τ_{nti} can be taken to be $I(|i-t|\leq\kappa_n)$, in which case $\lambda_{nij}=\big(1-|i-j|/(2\kappa_n+1)\big)I(|i-j|\leq 2\kappa_n)$ if $\kappa_n<i,j<n-\kappa_n$,[14] which are effectively the Newey–West weights.[15] A similar scheme for spatial data that are located on a two–dimensional lattice yields a weight — for observations located at (i,t) and (j,s) respectively — equal to

$$\lambda_{nitjs}=\frac{2\big(\min(|i-j|,|t-s|)+1\big)\big(\nu_{nitjs}+1\big)+\nu_{nitjs}^2-I(\nu_{nitjs}\text{ odd})}{(2\kappa_n+1)^2+1}I(\nu_{nitjs}\geq 0), \qquad (27)$$

where $\nu_{nitjs}=2\kappa_n-(|i-j|+|t-s|)$, provided that both observations are sufficiently far from the edge of the lattice. From (27) it follows that for fixed i,j,t,s, $\lambda_{nitjs}\to 1$ as $\kappa_n\to\infty$ since $\nu_{nitjs}/(2\kappa_n)\to 1$ as $\kappa_n\to\infty$.

Many schemes for the choice of the τ_{nti}–weights are possible. The simplest schemes have all weights equal to either zero or one; we use such a scheme in our application. But many other configurations are possible, including ones with negative weights.

3 The Application

We apply our technique to a panel of Canadian copper mines. The discrete choice in that application is to operate a property or to let it lie idle. In this section, we discuss the predictions from a theoretical real–options model, and we describe the copper mining industry and the data.

3.1 The Theory

Ever since the early 1980s, financial economists have realized that ownership of real assets has much in common with ownership of financial assets, and that the techniques that were developed to price and manage the former could be used to price and manage the latter.[16] In this subsection, we develop a reduced–form discrete–choice equation that embodies many of the predictions of the theory of real options. The equation that we specify is similar to the one that is used in Moel and Tufano (2002), which is itself based on the theoretical model of Brennan and Schwartz (1985). However, although we focus on a real–options model, our empirical specification allows us to discriminate among competing theories of operating decisions.

We consider a competitive industry, copper mining, and analyze flexible operating rules. Each mining property can be in one of two states — active or inactive — at the beginning of each period, and the manager must decide whether to operate the property or not. Since there are fixed costs associated with opening and closing, mines do not open (close) as soon as profits are expected to be positive (negative). Instead, there are price thresholds that must be crossed before an action is taken. Furthermore, the threshold that triggers opening is strictly higher than the one that triggers

[13] Note that $\hat{V}_n=\tilde{\psi}_n'\Lambda_n\tilde{\psi}_n$ with ψ_n a vector with i–th element $g_{ni}-\bar{g}_n$.

[14] For i,j closer to 1 or n the formula is somewhat more complex.

[15] The Newey–West weights would be $\big(1-|i-j|/(2\kappa_n+1)\big)I(|i-j|\leq 2\kappa_n)/(1-|i-j|/n)$.

[16] See, e.g., Tourinho (1979), Brennan and Schwartz (1985), and McDonald and Siegal (1985) for early applications of the theory of financial options to real assets. For a more recent and comprehensive treatment, see Dixit and Pindyck (1994).

closing. In the dynamic discrete–choice literature, the triggers are called S, s thresholds, with $S > s$. In our application, both thresholds are functions of the current state.

Under simple assumptions on the Data Generating Process for price,[17] it is possible to solve a real–options model to obtain a number of predictions concerning the determinants of the S, s thresholds and the optimal decision rules. Predictions that cause the thresholds to move in opposite directions or with different magnitudes are state dependent, whereas those that cause them to move in a similar fashion are not. The following is a list of predictions:

i) *Price*: High prices make it more likely that mines will be active, regardless of the prior state. In other words, high prices lower both thresholds.

ii) *Operating costs*: Mines with higher average variable costs are less likely to be active, regardless of the prior state. Furthermore, if fixed costs are not sunk (i.e., if they are not incurred when not operating), mines with higher fixed costs are also less likely to be active. In other words, high costs raise both thresholds.

iii) *Reserves*: Large reserves make it more likely that a mine will be active, regardless of the prior state. In other words, large reserves lower both thresholds.

iv) *Prior state*: A mine that is active (inactive) is more likely to remain active (inactive).

v) *Capacity*: The effect of capacity is less clear. However, if size is a reasonable proxy for opening and closing costs, a large capacity will delay actions. In other words, large mines that are active (inactive) will be more likely to remain active (inactive). A large capacity then raises the upper threshold, lowers the lower threshold, and widens the region of no change.

vi) *Price volatility*: High volatility delays actions. In particular, a mine that is active (inactive) is more likely to remain active (inactive) when prices are volatile. This means that high volatility widens the region between the thresholds (the region of inactivity).

Most of the predictions that are on our list would emerge from many other models of optimal operating decisions. For example, it is hard to think of a model that does not yield the first three predictions. Furthermore, any model with fixed opening and closing costs is apt to yield predictions iv) and v).

Only the final prediction distinguishes the real–options model from other models of optimal operating decisions. We therefore discuss that prediction in greater depth. With a real–options model, an increase in uncertainty increases the option value of delay. Managers are therefore less likely to open a currently inactive mine or close one that is active. With some competing investment models, such as those based on discounted–cash–flow (DCF), an increase in uncertainty has no effect. With others, an increase in volatility lowers the value of a project and makes operation less likely, regardless of the prior state. This is the case, for example, when investors are risk averse and face a tradeoff between the mean and the variance of returns. Prediction vi) therefore allows us to distinguish among theories.

3.2 The Industry

We apply our technique to a panel of Canadian copper mines. There is considerable variation in the size of Canadian mines. Nevertheless, we assume that all decision makers are price takers.

[17] The standard assumption is that price follows a geometric Brownian motion.

Indeed, the copper market is worldwide, and even the largest Canadian companies account for only a relatively small fraction of world copper production.[18]

There are a number of phases in the production of copper. In particular, ore is mined and concentrated at the mine site and then sent to smelters and refineries that produce metal. Most mining firms are vertically integrated into all phases of production. However, some smelting and refining of Canadian ores is performed by independent companies. When this occurs, however, it is customary for the mining company to pay a fixed charge per unit of metal and to continue to be the residual claimant (in other words, the mining company bears the price risk). This means that it is possible to know how much metal a mine produces each year and at what cost. Furthermore, it means that the price of refined copper is the relevant price, and that a measure of variation in that price is an appropriate measure of volatility.

We consider only developed properties. In other words, there are no mines that opened *de novo* in our sample. In fact, no Canadian copper mines opened *de novo* during the sample period. The copper industry, however, is highly cyclical, and mines frequently close when times are bad and reopen when conditions improve. Furthermore, most of the mines in our sample were inactive in some sample years, but no mine was inactive in all but one year.

Some mines are surface deposits whereas others are underground veins. There are therefore two mining technologies, open pit and underground. Furthermore, most open–pit mines are in the province of British Columbia, whereas most underground mines are in Quebec.

Spatial dependence can be due to a number of factors. Mines that are located close to one another are apt to have similar unobserved attributes that can lead to spatial error dependence. For example, although many supply and demand shocks are aggregate (e.g., those associated with energy prices and industrial production), some cost shocks are local. Regional cost variation can be due to, for example, common geological features and local input prices. When those factors are not observed, spatial dependence is incorporated into the errors.

3.3 The Data

The data that are used in the application pertain to a panel of 21 Canadian copper mines that were observed over a 14–year period, 1980 – 1993. The 21 mines were chosen to satisfy two criteria. First, their principal commodity had to be copper, and second, they had to be active during some portion of the 1980–1993 period. All mines that satisfy those criteria are included in the sample. Figures 2 and 3 contain maps of mine locations.

Most of the data, which are annual, are described in Slade (2001). A detailed description of the variables that are used in the current application can be found in the data appendix (C). A shorter description of each variable follows.

Mine data.

The variables that vary by mine and year are:

ACTIVE: A binary variable that equals 1 if the mine is active in the current period and zero otherwise.

RES: Reserves of ore remaining in the mine.

CAP: Mine capacity (in units of ore).

QCU: Output of metal.

COST: Real cost per unit of metal (available only when the mine is active).

[18]The Hirschman/Hirfindalh index for western–world copper mining averages 400 (see Slade and Thille 2005).

DOPEN: A binary variable that equals one if the technology is open pit and zero otherwise.[19]

The variables that vary only by mine are:
LAT and LONG: The spatial coordinates of the mine.
FCOST and MCOST: Fixed and marginal costs. The variable COST measures unit or average total cost. We generated a fixed and a marginal cost for each mine, FCOST_i and MCOST_i, using the following equation and the observations for which the mine was active,

$$TCOST_{it} \equiv COST_{it} \times QCU_{it} = FCOST_i + MCOST_i \times QCU_{it} + \epsilon_{it}, \tag{28}$$

where ϵ is a zero–mean random variable. Estimated fixed and marginal costs for mine i, which are the parameters in equation (28), are denoted $\hat{FCOST}_i$ and $\hat{MCOST}_i$.

Price data
The price and volatility data are common to all mines.
CPR: Real copper price (London Metal Exchange cash–settlement price).
SIGRR: Yearly volatility of real monthly returns,

$$SIGRR = SD(\frac{CPR - CPR_{-1}}{CPR_{-1}}), \tag{29}$$

where $SD(.)$ denotes the standard deviation, and -1 indicates the previous month.

There are 294 observations, 185 or 63% of which are active and 109 or 37% are inactive. Table 1 contains summary statistics for the dependent variable as well as for each variable that is used as a regressor. The first two columns show means and standard deviations for the entire sample, whereas subsequent columns contain separate summary statistics for active and inactive mines. This table suggests that mines are more apt to be active if prices, reserves, and capacity are high and volatility is low. Furthermore, on average, active mines have lower marginal costs. However, they have higher fixed costs, which is due to the fact that large mines are more apt to be active. Finally, open–pit mines are less likely to close than underground mines, which could be an indication of lower operating costs.

4 Results

The discrete–choice equations explain operating decisions for each mine in each year. It is assumed that those decisions are made at the beginning of the period using the information that is available at that time, Υ_t, which in practice is a set of excluded variables plus the set of $t - j$ regressors with $j \geq 1$.[20] Since expectations are formed rationally, expected and realized values of period–t explanatory variables differ by an error that is orthogonal to Υ_t. Lagged regressors are therefore appropriate instruments for the period–t variables that appear in the discrete–choice equations.

The discrete–choice models that we estimate are all special cases of equation (1), which, as discussed in that section, nests several models. The x vector that appears in that equation contains the variables that are shown in table 1.

All equations are estimated by one–step GMM with lagged (and/or current for the static models without fixed effects) regressors used as instruments; the first three lags are used for all years except

[19] Some mines have both a shaft and a pit, and the section of the mine that is active can vary.
[20] The data appendix discusses the excluded as well as the included variables.

for the first two, for which fewer lags are available. Specifications differ according to whether the equations are static or dynamic and whether they include fixed effects (denoted FE in the tables). Tables 2 and 3 contain ordinary probit models, albeit that they were estimated using GMM with the regressors as instruments. Since the number of restrictions is then equal to the number of unknowns, no weighting matrix needs to be chosen. For tables 4 and 5, excluded variables and lagged regressors were used as instruments, the number of instruments exceeds the number of unknown coefficients, and the Λ_n–matrix used in generating the GMM weighting matrix is the identity matrix. This procedure, which corrects for heteroskedasticity but not for spatial dependence, generates consistent estimates and, absent dependence, also appropriate standard errors. For tables 6 and 7, the same instruments were chosen as for tables 4 and 5, but the procedure of section 2.4 was used to create Λ_n. We set $\tau_{nit,js} = 1$ if mine j is among mine i's two closest neighbors[21] and $|t - s| \leq 1$, otherwise it is set to zero. We have also experimented with varying the number of neighbors but qualitatively the results do not change.

If the errors are heteroskedastic, not all specifications are consistent. However, we report models that range from the simplest static probit to the full specification so that sensitivity can be assessed.

4.1 Ordinary Probit Estimates

Table 2 contains static ordinary probit estimates of the operating rule. The dependent variable equals one when the mine is active and zero otherwise. The GMM moment conditions make use of the fact that the generalized errors from the probit should be orthogonal to the instruments (see Pinkse and Slade 1998).[22]

Consider first the specifications without fixed effects (numbers 1–6). They imply that mines are more apt to operate when price is high, volatility is low, the mine is large, and marginal cost is small. The effect of the fixed cost, in contrast, is not significant at conventional levels. In addition, mines with larger reserves are more likely to be active, except in specifications that include capacity, in which case the effect of reserves is not significant. This finding is probably due to the fact that RES and CAP are highly correlated. Finally, the effect of the type of mining technology, open pit or underground, is not significant.

With the ordinary probits, the mine fixed effects are included in the latent–variable equation rather than in the observed binary–choice equation, and no differencing is required. Since the number of time series observations per mine is 14, doing so may be reasonable in this instance. Specifications 7–9 in table 2 show that the inclusion of fixed effects increases the magnitude and the significance of the coefficients of the price and volatility variables. However, the coefficients of the other explanatory variables become insignificant. The fall in significance is due to the fact that, with fixed effects, identification is achieved through time–series variation, and there is much less variation in the time–series for reserves and capacity than in the cross–section.

The last column in table 2 contains the normalized success index, which is a measure of goodness of fit.[23] That index shows that performance is poor for the basic model (# 2) and that the addition of capacity has the biggest effect on performance. In what follows, we therefore emphasize variants of specification # 3, which includes capacity. The table also shows that the addition of fixed effects

[21] In Euclidean distance, see Pinkse, Slade and Brett (2002) for other notions of distance.

[22] When the generated variables, $\hat{FCOST}$ and $\hat{MCOST}$, are included, the standard errors are apt to be underestimated.

[23] See Hensher and Johnson (1981, p. 54) for a justification for this index, which is discussed in appendix D.

improves predictive power very substantially. We therefore also report variants of specification # 9.

The specifications in table 2 explain the probability that a mine will operate (not operate) in a given period. Equation (1), in contrast, which is a form of switching–regression with known switch points, explains the transition probabilities. In other words, it determines the probability of being in state k in period t, conditional on having been in state j in period $t-1$. The specifications in table 3 are more flexible in that some of the coefficients are allowed to depend on the regime (i.e., on the lagged dependent variable).

Recall that the effects of both volatility and capacity were hypothesized to depend on the previous state. Indeed, the real–options model predicts that increased uncertainty (higher volatility) causes the option value of a property to rise and delays investment. Furthermore, if a large capacity is a reasonable proxy for opening and closing costs, a large capacity should also delay closings and reopenings. This means that the coefficients of SIGRR and CAP should be positive when the mine was previously active but negative when it was previously inactive.

Table 3 contains specifications in which the effects of volatility and capacity are allowed to vary by state. It shows, however, that in no case does the sign of an estimated coefficient depend on the previous state and that this conclusion is independent of whether the specification includes mine fixed effects. Both the magnitude and the significance of the coefficients of SIGRR, however, are state dependent. In particular, the effect of volatility is much stronger when the mine was previously inactive. This means that increased volatility delays openings significantly but has a much smaller negative impact on closings.

The finding that the sign of the coefficient of SIGRR is negative regardless of the previous state is inconsistent with the real–options model. Instead, the dampening effect of volatility on operation is consistent with a more conventional mean/variance–utility model in which investors demand a higher mean when returns are more volatile. This result can be contrasted with that of Moel and Tufano (2002) who find evidence in favor of the real–options model in the gold–mining industry. In particular, their estimated coefficients are positive and negative respectively, as the real–options model predicts. Nevertheless, they find that the effect of volatility is not significant when the mine was previously closed.

The coefficients of the non–price variables in table 3 are not significant at conventional levels.[24] Moreover, comparing the estimates in tables 2 and 3, it is clear that significance tends to be lower overall when the model is dynamic. This is due to the fact that although, for example, there are 195 observations where a mine is active, a transition to a new state (e.g., to inactive from active) occurs for only 17 or 9% of them. Furthermore, there is little time–series variation in the mine–specific variables such as capacity and reserves.

Finally, the last column in table 3 shows that the normalized success index rises significantly when state dependence is considered (from, for example, 12% to 68% for specification # 3). Moreover, although the inclusion of fixed effects improves predicive power in table 3, in contrast to the results in table 2, the improvement is not dramatic. This difference is due to the fact that predictive power was already high.

4.2 Corrected Estimates

Correction for Heteroskedasticity

[24] This is also true in specifications that include the other explanatory variables that are shown in table 2.

Tables 4 and 5 contain the principal specifications from tables 2 and 3, respectively. However, the GMM estimation now uses the procedure of section 2.1. Specifically, we correct for heteroskedasticity but not for spatial dependence.

Consider table 4, which contains specifications that are not state dependent. One can see that the specifications without fixed effects have estimated coefficients that are typically less significant than those in table 2. Nevertheless, the signs and magnitudes of the coefficients, as well as the normalized success indices, are not very different across tables.[25]

When mine fixed effects are introduced, significance declines once again. The lack of significance with fixed effects is not surprising. Indeed, both the differencing procedure and the heteroskedasticity correction tend to reduce the number of significant coefficients. When they are combined, the reductions are compounded.

Table 5, like table 3, allows the coefficients of volatility and capacity to depend on the previous state. That table shows that, in general, significance is reduced compared to table 3. However, the magnitudes of the coefficients of volatility are considerably larger than those in table 3. Indeed, increased magnitude is observed with all specifications in table 5. It appears that modeling mine heterogeneity in the form of heteroskedastic errors allows one to uncover stronger volatility effects. Finally, correcting for heteroskedasticity also increases the predictive power of the model (the NSI rises from, for example, 68% to 79% for specification #3).

Spatial Correction

Tables 6 and 7 are comparable to tables 4 and 5 except that a correction for spatial and time series dependence was added. Specifically we use equation (26) with weights determined as discussed above. It is clear from the tables that, when both time–series and spatial dependence are modeled, the signs and magnitudes of the coefficients are similar to those obtained when only heteroskedasticity is corrected for. However, the coefficients are measured with greater accuracy in the full model (e.g., compare specification #6 in tables 5 and 7). Nevertheless, the predictive power of the spatial models in no better.

4.3 Transition Probabilities

The results from our state–dependent discrete–choice model are not supportive of the real–options model. A simple examination of the signs of the estimated coefficients in tables 3, 5, and 7 reveals this fact. The magnitudes of the coefficients in a probit switching regression, however, are more difficult to interpret directly. Here we examine magnitudes indirectly.

Table 8 illustrates how an increase in volatility affects operating decisions. Each part of that table is a 2×2 matrix of transition probabilities — the probability of being in state i conditional on having been in state j. The labels on the rows indicate the prior state, whereas those on the columns indicate the operating decision. The first half of the table is evaluated at the mean of the explanatory variables. It shows that 9% (7%) of active (inactive) mines close (open) in a typical period.[26]

The second half of the table differs from the first only in the value of our measure of price volatility, SIGRR. Instead of using mean volatility, we chose the mean plus one standard deviation, which we denote 'high volatility.' The table shows that the probability of operation declines in both

[25]The table does not show the normalized success index (NSI) for the specifications that have been differenced. Indeed, calculation of the NSI requires estimates of the fixed effects, which have been removed by differencing.

[26]Specification # 3 in table 3 was used to produce table 8.

states. However, the effect on mines that were previously closed is more dramatic, since, for those mines, the probability of opening when volatility is moderately high is virtually zero.

5 Conclusions

We have developed a one–step GMM (or continuous updating) estimator for dynamic discrete–choice models with endogenous regressors, fixed effects, and arbitrary patterns of spatial and time–series dependence. Furthermore, we have demonstrated that our estimator is consistent and asymptotically normal. That estimator is used to investigate closing and reopening decisions for a panel of Canadian copper mines. The exercise, which is cast in a real–options framework, is a two–state optimal–switching decision problem in which a mine can be either active or inactive, and the decision maker must decide whether to operate the mine or to let it lie idle. Our dynamic discrete–choice model allows us to use a Markov structure and to estimate Markov transition probabilities.

We find little support for the real–options model. Indeed, the data are more supportive of a more conventional model of investment or operating decisions in which the decision maker is risk averse and faces a tradeoff between the mean and the variance of returns.

Our one–step GMM estimator works well on simple models, both static and dynamic. However, when the model becomes more complex, for example, through the introduction of a combination of fixed effects and a weighting scheme with a large number of parameters, the estimated coefficients loose significance. Nevertheless, although significance tends to drop when more flexible specifications are estimated, the basic conclusions remain intact. Moreover, modeling state dependence improves the predictive performance of the model dramatically. Furthermore, modeling mine heterogeneity in the form of heteroskedastic errors not only improves performance but also uncovers larger effects of volatility.

A Continuous Updating

PLEASE NOTE: Throughout the appendix we indicate when convenient the assumption, lemma, equation or well–known theorem an (in)equality is based on above that (in)equality; an assumption is indicated by its letter, a common theorem by its name, a lemma by the letter L followed by its number and an equation by its equation number in brackets.

A.1 Generic Uniform Convergence

Lemma 1 is a self–contained generic uniform convergence result. There are many similar results available, see e.g. Andrews (1987, 1992) and Pötscher and Prucha (1989); we use this one simply because it is more convenient.

Lemma 1 *Let Θ be some compact convex set and Q_n some sequence of random functions which are differentiable with respect to $\theta \in \mathbb{R}^d$. Let for any $p > 0$,*

$$\mu_{np} = \max_{\theta\in\Theta}\big(E|Q_n(\theta)|^p\big)^{1/p}, \tag{30}$$

and let Δ_n be such that

$$\max_{\theta\in\Theta}\left|\left|\frac{\partial Q_n}{\partial\theta}(\theta)\right|\right| = O_p(\Delta_n). \tag{31}$$

Then,

$$\max_{\theta\in\Theta}|Q_n(\theta)| = o_p\Big(\mu_{np}^{\frac{p}{p+d}}\Delta_n^{\frac{d}{p+d}}\mathcal{L}_n\Big), \tag{32}$$

Proof: Denote the convergence rate in (32) by ϵ_n. Let $[t]_+$ denote the smallest integer greater than or equal to t. Divide the parameter space Θ up into $\upsilon_n = [\mathcal{C}/\delta_n^d]_+$ (for some $0 < \mathcal{C} < \infty$) possibly overlapping regions Θ_i and let $\tilde{\theta}_i \in \Theta_i$ be such that for all $\theta \in \Theta_i$, $||\theta - \tilde{\theta}_i|| < \delta_n$. This can be done by the convexity and compactness of Θ. Note that

$$\begin{aligned}
P\left(\max_{\theta\in\Theta}|Q_n(\theta)| > 2\epsilon_n\right) &= P\left(\max_{i\le\upsilon_n}\max_{\theta\in\Theta_i}|Q_n(\theta)| > 2\epsilon_n\right), \\
&\le P\left(\max_{i\le\upsilon_n}\max_{\theta\in\Theta_i}|Q_n(\theta) - Q_n(\tilde{\theta}_i)| > \epsilon_n\right) + P\left(\max_{i\le\upsilon_n}|Q_n(\tilde{\theta}_i)| > \epsilon_n\right) \\
&\le P\left(\max_{\theta\in\Theta}\left|\left|\frac{\partial Q_n}{\partial\theta}(\theta)\right|\right|\delta_n > \epsilon_n\right) + \sum_{i=1}^{\upsilon_n} P\left(|Q_n(\tilde{\theta}_i)| > \epsilon_n\right) \\
&\le P\left(\max_{\theta\in\Theta}\left|\left|\frac{\partial Q_n}{\partial\theta}(\theta)\right|\right| > \epsilon_n/\delta_n\right) + \upsilon_n\max_{\theta\in\Theta}P\left(|Q_n(\theta)| > \epsilon_n\right) \\
&\le P\left(\Delta_n^{-1}\max_{\theta\in\Theta}\left|\left|\frac{\partial Q_n}{\partial\theta}(\theta)\right|\right| > \epsilon_n\delta_n^{-1}\Delta_n^{-1}\right) + \upsilon_n\epsilon_n^{-p}\mu_{np}^p,
\end{aligned} \tag{33}$$

by the Markov inequality. Now choose $\delta_n = (\mu_{np}/\Delta_n)^{p/(p+d)}$ such that

$$\epsilon_n\delta_n^{-1}\Delta_n^{-1} = \mathcal{L}_n \to \infty, \quad \upsilon_n\epsilon_n^{-p}\mu_{np}^p = \mathcal{C}/\mathcal{L}_n^p \to 0,$$

both as $n \to \infty$. The first RHS term in (33) is then $o(1)$ by (31). ✓.

If, for example, Q_n is an average of i.i.d. mean zero random functions of θ, then $\mu_{np} = n^{-1/2}$, $\Delta_n = 1$ and the convergence rate is $n^{-p/\left(2(p+d)\right)}\mathcal{L}_n$, where $\mathcal{L}_n$ can be *slowly varying*. If an infinite number of moments exists, then the uniform convergence rate is close to root–n; see e.g. Pollard (1984), theorem 37, for a result in this case.[27]

Lemma 1 can be iterated if more derivatives exist. In the i.i.d. average case a convergence rate of

$$\mu_{np}^{1-\left(d/(p+d)\right)^{\mathcal{D}+1}}\mathcal{L}_n,$$

with $\mathcal{D}$ the number of derivatives is then achievable.

In our paper, Q_n will not always be an average, unlike the results of e.g. Andrews (1987) and Pötscher and Prucha (1989), and certainly not one of stationary objects. What lemma 1 achieves is that it directly expresses the uniform convergence rate of Q_n in terms of convergence and/or divergence rates which are more easily determined. The uniform convergence rate is not necessarily sharp, but it suffices for our purposes.

Finally, the notation $\mathcal{L}_n$ in lemma 1 intentionally coincides with that of assumption I; since any sequence $\{\mathcal{L}_n\}$ in lemma 1 will do provided that it tends to ∞, it can always be taken identical to the one in assumption I.

A.2 Consistency

Here the conditions of theorem 3 apply.

Let $\tilde{g}_{ni}(\theta) = g_{ni}(\theta) - Eg_{ni}(\theta)$, $\tilde{g}_n(\theta) = \bar{g}_n(\theta) - g_n(\theta)$ and $c_n = n^{-1/(2+d)}A_n^{1/(2+d)}\mathcal{L}_n$.

Lemma 2

$$\max_{\theta\in\Theta}||\tilde{g}_n(\theta)|| = o_p(c_n). \tag{34}$$

Proof: We use lemma 1 with $Q_n = \tilde{g}_n$ and $p = 2$. We first determine μ_{n2}.

$$\max_{\theta\in\Theta}E||\tilde{g}_n||^2 = \max_{\theta\in\Theta}n^{-2}\sum_{i,j=1}^{n}E(\tilde{g}_{ni}'\tilde{g}_{nj}) = \max_{\theta\in\Theta}n^{-2}\sum_{i,j=1}^{n}\sum_{t=1}^{d_g}E(\tilde{g}_{nit}\tilde{g}_{njt})$$

$$\overset{(18)}{\leq} n^{-2}\sum_{i,j=1}^{n}\alpha_{nij}\sum_{t=1}^{d_g}\sqrt{\max_{\theta\in\Theta}E||\tilde{g}_{nit}||^2}\sqrt{\max_{\theta\in\Theta}E||\tilde{g}_{njt}||^2} \leq n^{-2}\sum_{i,j=1}^{n}\alpha_{nij}d_g\sqrt{\max_{\theta\in\Theta}E||\tilde{g}_{ni}||^2}\sqrt{\max_{\theta\in\Theta}E||\tilde{g}_{nj}||^2}$$

$$\overset{\text{Liapunov}}{\leq} n^{-2}\sum_{i,j=1}^{n}\alpha_{nij}d_g\left(\max_{\theta\in\Theta}E||\tilde{g}_{ni}||^4\right)^{1/4}\left(\max_{\theta\in\Theta}E||\tilde{g}_{nj}||^4\right)^{1/4} \overset{\text{E}}{\leq} d_g\chi_n^{1/2}n^{-2}\sum_{i,j=1}^{n}\alpha_{nij} \overset{(19)}{\leq} d_g\chi_n^{1/2}n^{-1}A_n.$$

So $\mu_{n2} = (d_g\chi_n^{1/2}n^{-1}A_n)^{1/2}$. Now Δ_n. Note that

$$E\left(\max_{\theta\in\Theta}\left|\left|n^{-1}\sum_{i=1}^{n}\frac{\partial\tilde{g}_{ni}}{\partial\theta}(\theta)\right|\right|\right) \leq n^{-1}\sum_{i=1}^{n}E\left(\max_{\theta\in\Theta}\left|\left|\frac{\partial\tilde{g}_{ni}}{\partial\theta}(\theta)\right|\right|\right) \leq \frac{2}{n}\sum_{i=1}^{n}E\left(\max_{\theta\in\Theta}\left|\left|\frac{\partial g_{ni}}{\partial\theta}(\theta)\right|\right|\right)$$

$$\overset{\text{Liapunov}}{\leq} \frac{2}{n}\sum_{i=1}^{n}\left(E\left(\max_{\theta\in\Theta}\left|\left|\frac{\partial g_{ni}}{\partial\theta}(\theta)\right|\right|^2\right)\right)^{1/2} \overset{\text{E}}{\leq} 2\chi_n^{1/2}.$$

[27] Pollard's result does not require the existence of derivatives.

So $\Delta_n = \chi_n^{1/2}$. Hence the uniform convergence rate is

$$o_p\big(\mu_{n2}^{2/(2+d)}\Delta_n^{d/(2+d)}\mathcal{L}_n\big) = o_p\Big(\big(\chi_n^{1/2}n^{-1}A_n\big)^{1/(2+d)}\chi_n^{d/\left(2(2+d)\right)}\mathcal{L}_n\Big) \overset{\text{E}}{=} o_p\big((A_n/n)^{1/(2+d)}\mathcal{L}_n\big) = o_p(c_n),$$

as stated. ✓

Let

$$\mathcal{S}_0(\theta) = n^{-1}\sum_{i,j=1}^n \lambda_{nij}, \quad \mathcal{S}_1(\theta) = n^{-1}\sum_{i,j=1}^n \lambda_{nij}g_{ni}(\theta), \quad \mathcal{S}_2(\theta) = n^{-1}\sum_{i,j=1}^n \lambda_{nij}g_{ni}(\theta)g_{nj}'(\theta),$$

and $\mathcal{T}_r(\theta) = E\mathcal{S}_r(\theta)$ for $r = 0, 1, 2$.

Lemma 3 *For $r = 0, 1, 2$, $\max_{\theta\in\Theta}||\mathcal{S}_r(\theta) - \mathcal{T}_r(\theta)|| = o_p(c_n\Psi_n)$.*

Proof: We establish the proof for $r = 2$; the other two cases are similar and easier. For any $t, s = 1, \ldots, d_g$, let $\beta_{nij}(\theta) = \lambda_{nij}g_{nit}(\theta)g_{njs}(\theta) - E\big(\lambda_{nij}g_{nit}(\theta)g_{njs}(\theta)\big)$. We show convergence at the stated rate for each t, s, which implies that the established rate also applies to the norm. We use lemma 1 with $p = 2$. We first determine the value of μ_{n2}.

$$\begin{aligned}\max_{\theta\in\Theta} E\left(n^{-1}\sum_{i,j=1}^n \beta_{nij}(\theta)\right)^2 &= \max_{\theta\in\Theta} n^{-2}\sum_{i,j,i^*,j^*=1}^n E\big(\beta_{nij}(\theta)\beta_{ni^*j^*}(\theta)\big)\\ &\overset{(18)}{\leq} \max_{\theta\in\Theta} n^{-2}\sum_{i,j,i^*,j^*=1}^n \sqrt{E\beta_{nij}^2(\theta)E\beta_{ni^*j^*}^2(\theta)}(\alpha_{nii^*} + \alpha_{nij^*} + \alpha_{nji^*} + \alpha_{njj^*}). \quad (35)\end{aligned}$$

Now, by the law of iterated expectations,

$$\begin{aligned}\max_{\theta\in\Theta} E\beta_{nij}^2(\theta) &= \max_{\theta\in\Theta} E\Big(E\big(g_{nit}^2(\theta)g_{njs}^2(\theta)|\Lambda_n\big)\lambda_{nij}^2\Big)\\ &\overset{\text{Schwarz}}{\leq} \max_{\theta\in\Theta} E\left(\Big(E\big(g_{nit}^4(\theta)|\Lambda_n\big)\Big)^{1/2}\Big(E\big(g_{njs}^4(\theta)|\Lambda_n\big)\Big)^{1/2}\lambda_{nij}^2\right) \overset{\text{E}}{\leq} \chi_n E\lambda_{nij}^2. \quad (36)\end{aligned}$$

Thus,

$$\begin{aligned}\max_{\theta\in\Theta} n^{-2}\sum_{i,j,i^*,j^*=1}^n \sqrt{E\beta_{nij}^2(\theta)E\beta_{ni^*j^*}^2(\theta)}\alpha_{nii^*} &\overset{(36)}{\leq} n^{-2}\chi_n\sum_{i,i^*=1}^n \alpha_{nii^*}\sum_{j=1}^n\sqrt{E\lambda_{nij}^2}\sum_{j^*=1}^n\sqrt{E\lambda_{ni^*j^*}^2}\\ &\overset{(20)}{\leq} n^{-2}\chi_n\Psi_n^2\sum_{i,i^*=1}^n \alpha_{nii^*} \overset{(19)}{\leq} n^{-1}\chi_n\Psi_n^2A_n. \quad (37)\end{aligned}$$

Repeat the steps in (37) with $\alpha_{nij^*}, \alpha_{nji^*}, \alpha_{njj^*}$ in lieu of α_{nii^*} to establish that (35) is $O(n^{-1}\chi_n\Psi_n^2A_n)$. Since $\chi_n = O(1)$ by assumption E, μ_{n2} in lemma 1 is $O(n^{-1/2}\Psi_nA_n^{1/2})$.

We now determine the value of Δ_n in lemma 1. Note first that

$$\begin{aligned}&E\max_{\theta\in\Theta}\left|\left|\frac{\partial\beta_{nij}}{\partial\theta}(\theta)\right|\right|\\ &= E\max_{\theta\in\Theta}\left|\left|\lambda_{nij}\frac{\partial g_{nit}}{\partial\theta}(\theta)g_{njs}(\theta) - E\Big(\lambda_{nij}\frac{\partial g_{nit}}{\partial\theta}(\theta)g_{njs}(\theta)\Big) + \lambda_{nij}g_{nit}(\theta)\frac{\partial g_{njs}}{\partial\theta}(\theta) - E\Big(\lambda_{nij}g_{nit}(\theta)\frac{\partial g_{njs}}{\partial\theta}(\theta)\Big)\right|\right|\\ &\overset{\text{triangle}}{\leq} 2E\max_{\theta\in\Theta}\left|\left|\lambda_{nij}\frac{\partial g_{nit}}{\partial\theta}(\theta)g_{njs}(\theta)\right|\right| + 2E\max_{\theta\in\Theta}\left|\left|\lambda_{nij}g_{nit}(\theta)\frac{\partial g_{njs}}{\partial\theta}(\theta)\right|\right|. \quad (38)\end{aligned}$$

We only deal with the first RHS term in (38); the second RHS term follows similarly. We have

$$E\max_{\theta\in\Theta}\Big|\Big|\lambda_{nij}\frac{\partial g_{nit}}{\partial\theta}(\theta)g_{njs}(\theta)\Big|\Big| \overset{\text{Schwarz}}{\leq} \sqrt{E\Big(\lambda_{nij}^2\max_{\theta\in\Theta}g_{njs}^2(\theta)\Big)}\sqrt{E\max_{\theta\in\Theta}\Big|\Big|\frac{\partial g_{nit}}{\partial\theta}(\theta)\Big|\Big|^2}$$
$$\overset{(17)}{\leq} \chi_n^{1/2}\left(E\Big(E\big(\max_{\theta\in\Theta}g_{njs}^2(\theta)|\Lambda_n\big)\lambda_{nij}^2\Big)\right)^{1/2} \overset{\text{Liapunov}}{\leq} \chi_n^{1/2}\left(E\Big(\sqrt{E\big(\max_{\theta\in\Theta}g_{njs}^4(\theta)|\Lambda_n\big)}\lambda_{nij}^2\Big)\right)^{1/2}$$
$$\overset{(16)}{\leq} \chi_n^{3/4}\sqrt{E\lambda_{nij}^2}, \quad (39)$$

Use (39) in (38) to obtain that

$$E\left(\max_{\theta\in\Theta}\Big|\Big|n^{-1}\sum_{i,j=1}^n\frac{\partial\beta_{nij}}{\partial\theta}(\theta)\Big|\Big|\right) \leq n^{-1}\sum_{i,j=1}^n E\max_{\theta\in\Theta}\Big|\Big|\frac{\partial\beta_{nij}}{\partial\theta}(\theta)\Big|\Big| \leq 4\chi_n^{3/4}n^{-1}\sum_{i,j=1}^n E\sqrt{E\lambda_{nij}^2} \leq 4\chi_n^{3/4}\Psi_n.$$

Since $\chi_n = O(1)$ by assumption E, we can choose $\Delta_n = \Psi_n$ in lemma 1, and hence

$$\max_{\theta\in\Theta}\Big|\Big|n^{-1}\sum_{i,j=1}^n\beta_{nij}(\theta)\Big|\Big| = o_p\big(\mu_{n2}^{2/(2+d)}\Delta_n^{d/(2+d)}\mathcal{L}_n\big) = o_p\Big(\big(n^{-1/2}\Psi_nA_n^{1/2}\big)^{2/(2+d)}\Psi_n^{d/(2+d)}\mathcal{L}_n\Big) = o_p(c_n\Psi_n),$$

as stated. ✓.

Lemma 4 $\max_{\theta\in\Theta}||\hat{V}_n(\theta) - V_n(\theta)|| = o_p(c_n\Psi_n)$.
Proof: Note that (omitting the θ–arguments and recalling that $\tilde{g}_n = \bar{g}_n - g_n$),

$$\hat{V}_n = \mathcal{S}_2 - \bar{g}_n\mathcal{S}_1' - \mathcal{S}_1\bar{g}_n' + \mathcal{S}_0\bar{g}_n\bar{g}_n', \quad V_n = \mathcal{T}_2 - g_n\mathcal{T}_1' - \mathcal{T}_1g_n' + \mathcal{T}_0g_ng_n',$$

such that

$$\hat{V}_n - V_n = (\mathcal{S}_2 - \mathcal{T}_2) - \tilde{g}_n(\mathcal{S}_1 - \mathcal{T}_1)' - g_n(\mathcal{S}_1 - \mathcal{T}_1)' - \tilde{g}_n\mathcal{T}_1' - (\mathcal{S}_1 - \mathcal{T}_1)\tilde{g}_n' - (\mathcal{S}_1 - \mathcal{T}_1)g_n' - \mathcal{T}_1\tilde{g}_n'$$
$$+ (\mathcal{S}_0 - \mathcal{T}_0)(\bar{g}_n\bar{g}_n' - g_ng_n') + \mathcal{T}_0(\bar{g}_n\bar{g}_n' - gg') + (\mathcal{S}_0 - \mathcal{T}_0)g_ng_n'. \quad (40)$$

So if

$$c_n = O(1), \quad (41)$$
$$\max_{\theta\in\Theta}||\tilde{g}_n(\theta)|| = o_p(c_n), \quad (42)$$
$$\max_{\theta\in\Theta}||\mathcal{S}_r(\theta) - \mathcal{T}_r(\theta)|| = o_p(c_n\Psi_n), \quad r = 0,1,2, \quad (43)$$
$$\max_{\theta\in\Theta}||g_n(\theta)|| = O(1), \quad (44)$$
$$\max_{\theta\in\Theta}||\bar{g}_n(\theta)\bar{g}_n'(\theta) - g_n(\theta)g_n'(\theta)|| = o_p(c_n). \quad (45)$$
$$\max_{\theta\in\Theta}||\mathcal{T}_r(\theta)|| = O(\Psi_n), \quad r = 0,1, \quad (46)$$

then by (40)

$$\max_{\theta\in\Theta}||\hat{V}_n(\theta) - V_n(\theta)|| = o_p(c_n\Psi_n + c_n^2\Psi_n + c_n\Psi_n + c_n\Psi_n + c_n^2\Psi_n + c_n\Psi_n + c_n\Psi_n + c_n^2\Psi_n + c_n\Psi_n + c_n\Psi_n) = o_p(c_n\Psi_n).$$

First, (41) follows from assumption I since $c_n^{2+d} = n^{-1}A_n\mathcal{L}_n^{2+d} = o(\Psi_n^{-2(2+d)}) \overset{\text{H}}{=} o(1)$. Further, (42) and (43) follow from lemmas 2 and 3, respectively, and (44) holds since

$$\max_{\theta\in\Theta}||g_n(\theta)|| = \max_{\theta\in\Theta} E||n^{-1}\sum_{i=1}^n g_{ni}(\theta)|| \le n^{-1}\sum_{i=1}^n E\max_{\theta\in\Theta}||g_{ni}(\theta)||$$
$$\overset{\text{Liapunov}}{\le} \max_{i\le n}\big(E\max_{\theta\in\Theta}||g_{ni}(\theta)||^4\big)^{1/4} \overset{\text{E}}{\le} \chi_n^{1/4} = O(1).$$

Now (45). Using the expansion $\bar g_n\bar g_n' - g_ng_n' = (\bar g_n - g_n)(\bar g_n - g_n)' + g_n(\bar g_n - g_n)' + (\bar g_n - g_n)g_n' = \tilde g_n\tilde g_n' + g_n\tilde g_n' + \tilde g_n g_n'$ and the triangle inequality,

$$\max_{\theta\in\Theta}||\bar g_n(\theta)\bar g_n'(\theta) - g_n(\theta)g_n'(\theta)|| \le \max_{\theta\in\Theta}||\tilde g_n(\theta)||^2 + 2\max_{\theta\in\Theta}||g_n(\theta)||\max_{\theta\in\Theta}||\tilde g_n(\theta)|| \overset{(42),(44)}{=} o_p(c_n).$$

Finally (46). We show the case $r = 1$, the case $r = 0$ follows similarly.

$$\max_{\theta\in\Theta}||\mathcal{T}_1(\theta)|| \le n^{-1}\sum_{i,j=1}^n \max_{\theta\in\Theta} E\big|\big|\lambda_{nij}g_{ni}(\theta)\big|\big| = n^{-1}\sum_{i,j=1}^n \max_{\theta\in\Theta} E\big(E\big(||g_{ni}(\theta)||\,|\Lambda_n\big)\lambda_{nij}\big)$$
$$\overset{\text{Liapunov}}{\le} n^{-1}\sum_{i,j=1}^n \max_{\theta\in\Theta} E\left(\lambda_{nij}\Big(E\big(||g_{ni}(\theta)||^4|\Lambda_n\big)\Big)^{1/4}\right) \overset{\text{E}}{\le} n^{-1}\chi_n^{1/4}\sum_{i,j=1}^n E\lambda_{nij}$$
$$\overset{\text{Liapunov}}{\le} n^{-1}\chi_n^{1/4}\sum_{i,j=1}^n\sqrt{E\lambda_{nij}^2} \overset{(20)}{\le} \chi_n^{1/4}\Psi_n \overset{\text{E}}{=} O(\Psi_n), \quad \checkmark$$

Lemma 5 *(i)* $\min_{\theta\in\Theta}\mathcal{E}_{\min}\big(V_n(\theta)\big) \ge \rho_n^{-1}\zeta_n^{-1}$ *and (ii)* $\max_{\theta\in\Theta}\mathcal{E}_{\max}\big(V_n(\theta)\big) \le \Psi_n\zeta_n$.

Proof: First (i). Let $G_n(\theta)$ be a matrix with i–th row $n^{-1/2}\big(g_{ni}(\theta) - g_n(\theta)\big)'$. Then for all $v \ne 0$ and all $\theta\in\Theta$,

$$\frac{v'V_n(\theta)v}{v'v} = E\left(\frac{v'G_n'(\theta)\Lambda_nG_n(\theta)v}{v'v}\right) = E\left(\frac{v'G_n'(\theta)\Lambda_nG_n(\theta)v}{v'G_n'(\theta)G_n(\theta)v}\frac{v'G_n'(\theta)G_n(\theta)v}{v'v}\right)$$
$$\ge E\left(\mathcal{E}_{\min}(\Lambda_n)\frac{v'G_n'(\theta)G_n(\theta)v}{v'v}\right) \overset{\text{H}}{\ge} \rho_n^{-1}E\left(\frac{v'G_n'(\theta)G_n(\theta)v}{v'v}\right) \ge \rho_n^{-1}\frac{v'E\big(G_n'(\theta)G_n(\theta)\big)v}{v'v}$$
$$\ge \rho_n^{-1}n^{-1}\sum_{i=1}^n\frac{v'E\Big(\big(g_{ni}(\theta) - g_n(\theta)\big)\big(g_{ni}(\theta) - g_n(\theta)\big)'\Big)v}{v'v}$$
$$= \rho_n^{-1}n^{-1}\sum_{i=1}^n\frac{v'\Big(Vg_{ni}(\theta) + \big(Eg_{ni}(\theta) - g_n(\theta)\big)\big(Eg_{ni}(\theta) - g_n(\theta)\big)'\Big)v}{v'v}$$
$$\ge \rho_n^{-1}\frac{v'n^{-1}\sum_{i=1}^n Vg_{ni}(\theta)v}{v'v} \overset{\text{G}}{\ge} \rho_n^{-1}\mathcal{E}_{\min}\left(n^{-1}\sum_{i=1}^n Vg_{ni}(\theta)\right) \ge \rho_n^{-1}\zeta_n^{-1}.$$

Take the minimum over θ and note that the RHS does not depend on θ.

Now (ii). The proof is very similar to that of (i), albeit that we now use (20) to obtain an upper bound on $\mathcal{E}_{\max}(\Lambda_n)$. ✓

Lemma 6

$$\max_{\theta\in\Theta} |\hat{\Omega}_n(\theta) - \Omega_n(\theta)| = o_p(1). \tag{47}$$

Proof: Expand $\hat{\Omega}_n - \Omega_n$ as

$$\begin{aligned}\bar{g}_n'\hat{W}_n\bar{g}_n - g_n'W_ng_n &= (\bar{g}_n - g_n + g_n)'(\hat{W}_n - W_n + W_n)(\bar{g}_n - g_n + g_n) - g_n'W_ng_n \\ &= (\bar{g}_n - g_n)'(\hat{W}_n - W_n)(\bar{g}_n - g_n) + 2(\bar{g}_n - g_n)'(\hat{W}_n - W_n)g_n + g_n'(\hat{W}_n - W_n)g_n \\ &\quad + (\bar{g}_n - g_n)'W_n(\bar{g}_n - g_n) + 2(\bar{g}_n - g_n)'W_ng_n.\end{aligned}$$

Since $\Psi_n^{-1} = O(1)$ by assumption H it hence suffices to establish that

$$\max_{\theta\in\Theta} ||\bar{g}_n(\theta) - g_n(\theta)|| = o_p(\Psi_n^{-1}), \tag{48}$$

$$\max_{\theta\in\Theta} ||\hat{W}_n(\theta) - W_n(\theta)|| = o_p(1), \tag{49}$$

$$\max_{\theta\in\Theta} |g_n(\theta)| = O(1), \tag{50}$$

$$\max_{\theta\in\Theta} ||W_n(\theta)|| = O(\Psi_n). \tag{51}$$

(48) follows from (42) and assumption I; (50) follows from (44).

Now (51). Note that since W_n is positive definite,

$$\max_{\theta\in\Theta} ||W_n(\theta)|| = \max_{\theta\in\Theta} \mathcal{E}_{\max}\big(W_n(\theta)\big) = \frac{\Psi_n}{\min_{\theta\in\Theta} \mathcal{E}_{\min}\big(V_n(\theta)\big)} \overset{\text{L5}}{\leq} \Psi_n\rho_n\zeta_n = O(\Psi_n), \tag{52}$$

by assumptions G and H.

Finally (49). Note that

$$\begin{aligned}\max_{\theta\in\Theta} ||\hat{W}_n - W_n|| &= \max_{\theta\in\Theta} \big|\big|(\hat{W}_n - W_n)(W_n^{-1} - \hat{W}_n^{-1})W_n + W_n(W_n^{-1} - \hat{W}_n^{-1})W_n\big|\big| \\ &\overset{\text{triangle}}{\leq} \max_{\theta\in\Theta} \big|\big|\Psi_n^{-1}(\hat{W}_n - W_n)(V_n - \hat{V}_n)W_n\big|\big| + \max_{\theta\in\Theta} \big|\big|\Psi_n^{-1}W_n(V_n - \hat{V}_n)W_n\big|\big|. \end{aligned}\tag{53}$$

First, the second RHS term in (53). By (51), lemma 4 and assumption I,

$$\max_{\theta\in\Theta} \Big|\Big|\Psi_n^{-1}W_n(\theta)\big(V_n(\theta) - \hat{V}_n(\theta)\big)W_n(\theta)\Big|\Big| = O_p(\Psi_n^{-1}\Psi_nc_n\Psi_n\Psi_n) = O_p(c_n\Psi_n) = o_p(1).$$

The first RHS term in (53) is of smaller order than the LHS since

$$\begin{aligned}&\max_{\theta\in\Theta} \Big|\Big|\Psi_n^{-1}\big(\hat{W}_n(\theta) - W_n(\theta)\big)\big(V_n(\theta) - \hat{V}_n(\theta)\big)W_n(\theta)\Big|\Big| \\ &\qquad \leq \Psi_n^{-1} \max_{\theta\in\Theta} \big|\big|V_n(\theta) - \hat{V}_n(\theta)\big|\big| \max_{\theta\in\Theta} \big|\big|W_n(\theta)\big|\big| \max_{\theta\in\Theta} \big|\big|\hat{W}_n(\theta) - W_n(\theta)\big|\big| \\ &\qquad\qquad = O_p(c_n\Psi_n) \max_{\theta\in\Theta} \big|\big|\hat{W}_n(\theta) - W_n(\theta)\big|\big| = o_p(1) \max_{\theta\in\Theta} \big|\big|\hat{W}_n(\theta) - W_n(\theta)\big|\big|,\end{aligned}$$

again by (51), lemma 4 and assumption I. ✓

Proof of theorem 3: We show that $g^*(\hat{\theta}) = o_p(1)$ which by assumption F (continuity and $g^*(\theta) = 0 \Leftrightarrow \theta = \theta_0$) implies that $\hat{\theta} \xrightarrow{\mathcal{P}} \theta_0$. Now,

$$\big(g^*(\hat{\theta})\big)^2 \overset{\text{F}}{\leq} ||g_n(\hat{\theta})||^2 \leq \frac{\Psi_n g_n'(\hat{\theta}) V_n^{-1}(\hat{\theta}) g_n(\hat{\theta})}{\Psi_n \mathcal{E}_{\min}\big(V_n^{-1}(\hat{\theta})\big)} \leq \Psi_n^{-1} \Omega_n(\hat{\theta}) \mathcal{E}_{\max}\big(V_n(\hat{\theta})\big) \overset{\text{L5}}{\leq} \zeta_n \Omega_n(\hat{\theta}).$$

Since $\zeta_n = O(1)$ by assumption G, $\Omega_n(\hat{\theta}) \xrightarrow{\mathcal{P}} 0$ implies $g^*(\hat{\theta}) \xrightarrow{\mathcal{P}} 0$. So it suffices to show that $\Omega_n(\hat{\theta}) \xrightarrow{\mathcal{P}} 0$. Finally, noting that $\Omega_n(\theta_0) = 0$ and $\Omega_n(\hat{\theta}) \geq 0$,

$$0 \leq \Omega_n(\hat{\theta}) = \big(\Omega_n(\hat{\theta}) - \hat{\Omega}_n(\hat{\theta})\big) + \big(\hat{\Omega}_n(\hat{\theta}) - \hat{\Omega}_n(\theta_0)\big) + \big(\hat{\Omega}_n(\theta_0) - \Omega_n(\theta_0)\big) \leq o_p(1) + 0 + o_p(1) = o_p(1). \quad \checkmark$$

A.3 Asymptotic Normality

Here the assumptions of theorem 4 apply.

Lemma 7

$$\sqrt{n}\bar{g}_n(\theta_0) \xrightarrow{\mathcal{D}} N(0, V_0). \tag{54}$$

Proof: By assumption K, the conditions of theorem 2 are satisfied and hence

$$\Big(V_n^*(\theta_0)\Big)^{-1/2} \sqrt{n}\bar{g}_n(\theta_0) \xrightarrow{\mathcal{D}} N(0, I).$$

Since $V_0 = \lim_{n\to\infty} V_n^*(\theta_0)$, the stated result holds. ✓

Lemma 8 *For any consistent estimator $\hat{\theta}^*$ of θ_0,*

$$\frac{\partial \bar{g}_n}{\partial \theta'}(\hat{\theta}^*) - T_0 = o_p(1). \tag{55}$$

Proof: Write the LHS in (55) as

$$\left(\frac{\partial \bar{g}_n}{\partial \theta'}(\hat{\theta}^*) - \frac{\partial \bar{g}_n}{\partial \theta'}(\theta_0)\right) + \left(\frac{\partial \bar{g}_n}{\partial \theta'}(\theta_0) - \frac{\partial g_n}{\partial \theta'}(\theta_0)\right) + \left(\frac{\partial g_n}{\partial \theta'}(\theta_0) - T_0\right). \tag{56}$$

The third term in (56) is $o(1)$ by (24) and the first term is $o_p(1)$ by assumption M). Finally, the second term is also $o_p(1)$ since for any $t = 1, \ldots, d$,

$$E\left|\left|\frac{\partial \bar{g}_n}{\partial \theta_t}(\theta_0) - \frac{\partial g_n}{\partial \theta_t}(\theta_0)\right|\right|^2 \leq n^{-2} \sum_{i,j=1}^{n} E\left(\frac{\partial \tilde{g}_{ni}'}{\partial \theta_t}(\theta_0) \frac{\partial \tilde{g}_{nj}}{\partial \theta_t}(\theta_0)\right)$$

$$\overset{(25)}{\leq} n^{-2} \sum_{i,j=1}^{n} \sum_{t=1}^{d_g} \sqrt{E\left|\left|\frac{\partial \tilde{g}_{ni}}{\partial \theta_t}(\theta_0)\right|\right|^2} \sqrt{E\left|\left|\frac{\partial \tilde{g}_{nj}}{\partial \theta_t}(\theta_0)\right|\right|^2} \alpha_{nij} \overset{\text{E}}{\leq} d_g \chi_n n^{-2} \sum_{i,j=1}^{n} \alpha_{nij} \overset{(19)}{\leq} d_g \chi_n n^{-1} A_n \overset{\text{I}}{=} o(1). \quad \checkmark$$

In lemma 9 we establish a bound for the convergence rate of $\hat{\theta}$, which can then be used in subsequent results. The true convergence rate is later established to be root–n, as the theorem suggests.

Lemma 9 $||\hat{\theta} - \theta_0|| = O_p\big(n^{-1/2}\Psi_n^{1/2}\big)$.

Proof: By the mean value theorem for some $\hat{\theta}^*$ between θ_0 and $\hat{\theta}$,

$$\begin{aligned}\bar{g}_n(\hat{\theta}) &= \bar{g}_n(\hat{\theta}) - \bar{g}_n(\theta_0) + \bar{g}_n(\theta_0) \overset{\text{L7}}{=} \bar{g}_n(\hat{\theta}) - \bar{g}_n(\theta_0) + O_p(n^{-1/2}) \\ &= \frac{\partial \bar{g}_n}{\partial \theta'}(\hat{\theta}^*)(\hat{\theta} - \theta_0) + O_p(n^{-1/2}) = \left(\frac{\partial \bar{g}_n}{\partial \theta'}(\hat{\theta}^*) - T_0\right)(\hat{\theta} - \theta_0) + T_0(\hat{\theta} - \theta_0) + O_p(n^{-1/2}) \\ &\overset{\text{L8}}{=} \big(o_p(1) + T_0\big)(\hat{\theta} - \theta_0) + O_p(n^{-1/2}).\end{aligned}$$

Thus,

$$\begin{aligned}\hat{\theta} - \theta_0 &= \big(T_0' + o_p(1)\big)\big(T_0 + o_p(1)\big)^{-1}\big(T_0' + o_p(1)\big)\bar{g}_n(\hat{\theta}) + T_0' \times O_p(n^{-1/2}) \\ &\overset{\text{J}}{=} \big((T_0'T_0)^{-1}T_0' + o_p(1)\big)\bar{g}_n(\hat{\theta}) + O_p(n^{-1/2}) \overset{\text{J}}{=} O_p(1) \times \bar{g}_n(\hat{\theta}) + O_p(n^{-1/2}).\end{aligned}$$

So we need to show that $||\bar{g}_n(\hat{\theta})||^2 = O_p(n^{-1}\Psi_n)$. Thus,

$$||\bar{g}_n(\hat{\theta})||^2 \leq \frac{\bar{g}_n'(\hat{\theta})\hat{W}_n(\hat{\theta})\bar{g}_n(\hat{\theta})}{\mathcal{E}_{\min}\big(\hat{W}_n(\hat{\theta})\big)} = \frac{\hat{\Omega}_n(\hat{\theta})}{\mathcal{E}_{\min}\big(\hat{W}_n(\hat{\theta})\big)}. \tag{57}$$

We now work on each of the components of (57). Since $\hat{\theta}$ is a global minimizer of $\hat{\Omega}_n$ we have

$$0 \leq \hat{\Omega}_n(\hat{\theta}) \leq \hat{\Omega}_n(\theta_0) \leq ||\bar{g}_n(\theta_0)||^2 \mathcal{E}_{\max}\big(\hat{W}_n(\theta_0)\big) \overset{\text{L7}}{=} O_p(n^{-1}) \times \mathcal{E}_{\max}\big(\hat{W}_n(\theta_0)\big). \tag{58}$$

Further,

$$\mathcal{E}_{\max}\big(\hat{W}_n(\theta_0)\big) = ||\hat{W}_n(\theta_0)|| \overset{\text{triangle}}{\leq} ||\hat{W}_n(\theta_0) - W_n(\theta_0)|| + ||W_n(\theta_0)|| \overset{(49),(51)}{=} O_p(\Psi_n). \tag{59}$$

Finally,

$$\frac{1}{\mathcal{E}_{\min}\big(\hat{W}_n(\hat{\theta})\big)} = \mathcal{E}_{\max}\big(\hat{W}_n^{-1}(\hat{\theta})\big) = \Psi_n^{-1}\mathcal{E}_{\max}\big(\hat{V}_n(\hat{\theta})\big) \overset{\text{L5}}{\leq} \zeta_n \overset{\text{H}}{=} O_p(1). \tag{60}$$

Use (58)–(60) in (57). ✓

Lemma 10

$$V_n(\theta_0) - V_n^*(\theta_0) = o_p(1).$$

Proof: Since $g_n(\theta_0) = 0$,

$$V_n(\theta_0) = n^{-1}\sum_{i,j=1}^{n} E\big(\lambda_{nij} g_{ni}(\theta_0) g_{nj}'(\theta_0)\big).$$

Hence

$$V_n^*(\theta_0) - V_n(\theta_0) = n^{-1}\sum_{i,j=1}^{n} E\big((1 - \lambda_{nij}) g_{ni}(\theta_0) g_{nj}'(\theta_0)\big).$$

Use assumption L. ✓

Lemma 11 $V_n(\hat{\theta}) - V_n(\theta_0) = o_p(1)$.
Proof: By the mean value theorem for some $\hat{\theta}^*$ between $\hat{\theta}$ and θ_0,

$$V_n(\hat{\theta}) - V_n(\theta_0) = \sum_{s=1}^{d} \frac{\partial V_n}{\partial \theta_s}(\hat{\theta}^*)(\hat{\theta}_s - \theta_{0s}).$$

Since $\hat{\theta} - \theta_0 = O_p(n^{-1/2}\Psi_n^{1/2})$ by lemma 9, it suffices to show that for every $s = 1, \ldots, d$,

$$\max_{\theta\in\Theta} \left|\left| \frac{\partial V_n}{\partial \theta_s}(\theta) \right|\right| = o(n^{1/2}\Psi_n^{-1/2}). \tag{61}$$

Take any one such s. Note that

$$\begin{aligned}
\max_{\theta\in\Theta} \left|\left| \frac{\partial V_n}{\partial \theta_s}(\theta) \right|\right| &= \max_{\theta\in\Theta} \left|\left| n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij} \frac{\partial}{\partial \theta_s} \Big(\big(g_{ni}(\theta) - g_n(\theta)\big)\big(g_{nj}(\theta) - g_n(\theta)\big) \Big) \Big) \right|\right| \\
&\overset{\text{triangle}}{\leq} \max_{\theta\in\Theta} \left|\left| n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\theta) g'_{nj}(\theta) \Big) \right|\right| + \max_{\theta\in\Theta} \left|\left| n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij} g_{ni}(\theta) \frac{\partial g'_{nj}}{\partial \theta_s}(\theta) \Big) \right|\right| \qquad (62) \\
&\quad + \max_{\theta\in\Theta} \left|\left| n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\theta) \Big) g'_n(\theta) \right|\right| + \max_{\theta\in\Theta} \left|\left| g_n(\theta) n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij} \frac{\partial g'_{nj}}{\partial \theta_s}(\theta) \Big) \right|\right| \qquad (63) \\
&\quad + \max_{\theta\in\Theta} \left|\left| n^{-1} \sum_{i,j=1}^{n} E\big(\lambda_{nij} g_{ni}(\theta) \big) \frac{\partial g'_n}{\partial \theta_s}(\theta) \right|\right| + \max_{\theta\in\Theta} \left|\left| \frac{\partial g_n}{\partial \theta_s}(\theta) n^{-1} \sum_{i,j=1}^{n} E\big(\lambda_{nij} g'_{nj}(\theta) \big) \right|\right| \qquad (64) \\
&\quad + \max_{\theta\in\Theta} \left|\left| \left(\frac{\partial g_n}{\partial \theta_s}(\theta) g'_n(\theta) + g_n(\theta) \frac{\partial g'_n}{\partial \theta_s}(\theta) \right) n^{-1} \sum_{i,j=1}^{n} E\lambda_{nij} \right|\right|. \qquad (65)
\end{aligned}$$

Consider the first term in (62). We have

$$\begin{aligned}
\max_{\theta\in\Theta} \left|\left| n^{-1} \sum_{i,j=1}^{n} E\Big(\lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\theta) g'_{nj}(\theta) \Big) \right|\right| &\overset{\text{Schwarz}}{\leq} \max_{\theta\in\Theta} n^{-1} \sum_{i,j=1}^{n} \sqrt{E\big|\big|\lambda_{nij} g_{nj}(\theta)\big|\big|^2} \sqrt{E\left|\left| \frac{\partial g_{ni}}{\partial \theta_s}(\theta) \right|\right|^2} \\
&= \max_{\theta\in\Theta} n^{-1} \sum_{i,j=1}^{n} \sqrt{E\Big(\lambda_{nij}^2 E\big(\big|\big| g_{nj}(\theta) \big|\big|^2 | \Lambda_n \big) \Big)} \sqrt{E\left|\left| \frac{\partial g_{ni}}{\partial \theta_s}(\theta) \right|\right|^2} \\
&\overset{\text{Liapunov}}{\leq} \max_{\theta\in\Theta} n^{-1} \sum_{i,j=1}^{n} \sqrt{E\Big(\lambda_{nij}^2 \sqrt{E\big(\big|\big| g_{nj}(\theta) \big|\big|^4 | \Lambda_n \big)} \Big)} \sqrt{E\left|\left| \frac{\partial g_{ni}}{\partial \theta_s}(\theta) \right|\right|^2} \\
&\overset{\text{E}}{\leq} \chi_n^{3/4} n^{-1} \sum_{i,j=1}^{n} \sqrt{E\lambda_{nij}^2} \overset{(20)}{\leq} \chi_n^{3/4} \Psi_n \overset{\text{E}}{=} O(\Psi_n) \overset{\text{I}}{=} o(n^{1/2}\Psi_n^{-1/2}),
\end{aligned} \tag{66}$$

as required by (61). By symmetry the second term in (62) is also $O(\Psi_n)$. The remaining terms, i.e. those in (63)–(65) can be dealt with by a tedious repetition of steps similar to those in (66). ✓

Lemma 12

$$\Psi_n^{-1} \hat{W}_n(\hat{\theta}) \xrightarrow{\mathcal{P}} V_0^{-1}, \quad \Psi_n^{-1} \hat{W}_n(\theta_0) \xrightarrow{\mathcal{P}} V_0^{-1}. \tag{67}$$

Proof: Note that

$$\Psi_n^{-1}\hat{W}_n(\hat{\theta}) - V_0^{-1} = \Psi_n^{-1}\big(\hat{W}_n(\hat{\theta}) - W_n(\hat{\theta})\big) + \Psi_n^{-1}\big(W_n(\hat{\theta}) - W_n(\theta_0)\big) + \big(\Psi_n^{-1}W_n(\theta_0) - V_0^{-1}\big), \quad (68)$$

$$\Psi_n^{-1}\hat{W}_n(\theta_0) - V_0^{-1} = \Psi_n^{-1}\big(\hat{W}_n(\theta_0) - W_n(\theta_0)\big) + \big(\Psi_n^{-1}W_n(\theta_0) - V_0^{-1}\big). \quad (69)$$

We now show that all RHS terms in (68) and (69) are $o_p(1)$ or $o(1)$. The first RHS terms in both equations are $o_p(\Psi_n^{-1}) = o_p(1)$ by (49) and assumption J. Now, each of the last RHS terms is

$$V_n^{-1}(\theta_0) - V_0^{-1} = o(1), \quad (70)$$

since inversion is a continuous operation, $V_0 > 0$ by assumption J, and because

$$V_n(\theta_0) - V_0 = \big(V_n(\theta_0) - V_n^*(\theta_0)\big) + \big(V_n^*(\theta_0) - V_0\big) \overset{\text{L10,(24)}}{=} o(1).$$

Finally the second RHS term in (68). Note that

$$\begin{aligned}
\big|\big|\Psi_n^{-1}\big(W_n(\hat{\theta}) - W_n(\theta_0)\big)\big|\big| &= \big|\big|V_n^{-1}(\hat{\theta}) - V_n^{-1}(\theta_0)\big|\big| = \big|\big|V_n^{-1}(\hat{\theta})\big(V_n(\theta_0) - V_n(\hat{\theta})\big)V_n^{-1}(\theta_0)\big|\big| \\
&\overset{\text{triangle}}{\leq} \big|\big|\big(V_n^{-1}(\hat{\theta}) - V_n^{-1}(\theta_0)\big)\big(V_n(\theta_0) - V_n(\hat{\theta})\big)V_n^{-1}(\theta_0)\big|\big| + \big|\big|V_n^{-1}(\theta_0)\big(V_n(\theta_0) - V_n(\hat{\theta})\big)V_n^{-1}(\theta_0)\big|\big| \\
&\leq ||V_n^{-1}(\hat{\theta}) - V_n^{-1}(\theta_0)|| \cdot ||V_n(\theta_0) - V_n(\hat{\theta})|| \cdot ||V_n^{-1}(\theta_0)|| + ||V_n^{-1}(\theta_0)||^2 \cdot ||V_n(\theta_0) - V_n(\hat{\theta})|| \\
&\overset{\text{L11,(70)}}{=} \big|\big|\Psi_n^{-1}\big(W_n(\hat{\theta}) - W_n(\theta_0)\big)\big|\big| \times o_p(1) \times O(1) + O(1) \times o_p(1) = \big|\big|\Psi_n^{-1}\big(W_n(\hat{\theta}) - W_n(\theta_0)\big)\big|\big| \times o_p(1) + o_p(1).
\end{aligned}$$

Hence $\big|\big|\Psi_n^{-1}\big(W_n(\hat{\theta}) - W_n(\theta_0)\big)\big|\big| = o_p(1)$. ✓

Lemma 13

$$\bar{g}_n(\hat{\theta}) = O_p\big(n^{-1/2}\big). \quad (71)$$

Proof: Note that

$$\begin{aligned}
\mathcal{E}_{\min}(V_0^{-1})\big|\big|\bar{g}_n(\hat{\theta})\big|\big|^2 \leq \bar{g}_n'(\hat{\theta})V_0^{-1}\bar{g}_n(\hat{\theta}) &= \bar{g}_n'(\hat{\theta})\big(V_0^{-1} - \Psi_n^{-1}\hat{W}_n(\hat{\theta})\big)\bar{g}_n(\hat{\theta}) + \Psi_n^{-1}\hat{\Omega}_n(\hat{\theta}) \\
&\overset{\text{L12}}{\leq} o_p(1)\big|\big|\bar{g}_n(\hat{\theta})\big|\big|^2 + \Psi_n^{-1}\hat{\Omega}_n(\hat{\theta}). \quad (72)
\end{aligned}$$

Since V_0 is positive definite and finite by assumption J, (72) implies that the convergence rate of $||\bar{g}_n(\hat{\theta})||^2$ is no slower than that of $\Psi_n^{-1}\hat{\Omega}_n(\hat{\theta})$. Thus,

$$\begin{aligned}
\Psi_n^{-1}\hat{\Omega}_n(\hat{\theta}) \leq \Psi_n^{-1}\hat{\Omega}_n(\theta_0) &= \bar{g}_n'(\theta_0)\big(\Psi_n^{-1}\hat{W}_n(\theta_0) - V_0^{-1})\big)\bar{g}_n(\theta_0) + \bar{g}_n'(\theta_0)V_0^{-1}\bar{g}_n(\theta_0) \\
&\overset{\text{L12,J}}{\leq} \big(o_p(1) + \mathcal{E}_{\max}(V_0^{-1})\big)||\bar{g}_n(\theta_0)||^2 \overset{\text{L7}}{=} O_p(n^{-1}).
\end{aligned}$$

Hence $||\bar{g}_n(\hat{\theta})|| = O_p(n^{-1/2})$, as claimed. ✓

Lemma 14

$$\frac{\partial \hat{W}_n}{\partial \theta_s}(\hat{\theta}) = O_p(\Psi_n^2), \quad s = 1, \ldots, d. \quad (73)$$

Proof: Choose any $s = 1, \ldots, d$. Then

$$\frac{\partial \hat{W}_n}{\partial \theta_s}(\hat{\theta}) = -\hat{W}_n(\hat{\theta})\frac{\partial \hat{W}_n^{-1}}{\partial \theta_s}(\hat{\theta})\hat{W}_n(\hat{\theta}) = -\Psi_n^{-1}\hat{W}_n(\hat{\theta})\frac{\partial \hat{V}_n}{\partial \theta_s}(\hat{\theta})\hat{W}_n(\hat{\theta}) \overset{\text{L12}}{=} O_p(1) \times \frac{\partial \hat{V}_n}{\partial \theta_s}(\hat{\theta}) \times O_p(\Psi_n).$$

So it suffices to show that

$$\frac{\partial \hat{V}_n}{\partial \theta_s}(\hat{\theta}) = O_p(\Psi_n).$$

Now,

$$\begin{aligned}
\frac{\partial \hat{V}_n}{\partial \theta_s}(\hat{\theta}) &= \frac{\partial}{\partial \theta_s} n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \big(g_{ni}(\hat{\theta}) - \bar{g}_n(\hat{\theta})\big)\big(g_{nj}(\hat{\theta}) - \bar{g}_n(\hat{\theta})\big)' \\
&= n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \Big(\frac{\partial g_{ni}}{\partial \theta_s}(\hat{\theta}) g'_{nj}(\hat{\theta}) + g_{ni}(\hat{\theta}) \frac{\partial g_{nj'}}{\partial \theta_s}(\hat{\theta}) \Big) - \frac{\partial \bar{g}_n}{\partial \theta_s}(\hat{\theta}) n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} g'_{nj}(\hat{\theta}) - n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} g_{ni}(\hat{\theta}) \frac{\partial \bar{g}'_n}{\partial \theta_s}(\hat{\theta}) \\
&\quad - \bar{g}_n(\hat{\theta}) n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \frac{\partial g'_{nj}}{\partial \theta_s}(\hat{\theta}) - n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\hat{\theta}) \bar{g}'_n(\hat{\theta}) + \left(\frac{\partial \bar{g}_n}{\partial \theta_s}(\hat{\theta}) \bar{g}'_n(\hat{\theta}) + \bar{g}_n(\hat{\theta}) \frac{\partial \bar{g}'_n}{\partial \theta_s}(\hat{\theta}) \right) n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \\
&\overset{\text{L8,L13}}{=} n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \Big(\frac{\partial g_{ni}}{\partial \theta_s}(\hat{\theta}) g'_{nj}(\hat{\theta}) + g_{ni}(\hat{\theta}) \frac{\partial g_{nj'}}{\partial \theta_s}(\hat{\theta}) \Big) - O_p(1) \cdot n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} g'_{nj}(\hat{\theta}) - n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} g_{ni}(\hat{\theta}) \cdot O_p(1) \\
&\quad - O_p(n^{-1/2}) \cdot n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \frac{\partial g'_{nj}}{\partial \theta_s}(\hat{\theta}) - n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\hat{\theta}) \cdot O_p(n^{-1/2}) + O_p(n^{-1/2}) \cdot n^{-1} \sum_{i,j=1}^{n} \lambda_{nij}. \quad (74)
\end{aligned}$$

Now,

$$\begin{aligned}
E\Big|\Big| n^{-1} \sum_{i,j=1}^{n} \lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\hat{\theta}) g'_{nj}(\hat{\theta}) \Big|\Big| &\le n^{-1} \sum_{i,j=1}^{n} E \max_{\theta \in \Theta} \Big|\Big| \lambda_{nij} \frac{\partial g_{ni}}{\partial \theta_s}(\theta) g'_{nj}(\theta) \Big|\Big| \\
&\overset{\text{Schwarz}}{\le} n^{-1} \sum_{i,j=1}^{n} \sqrt{E\Big(\lambda_{nij}^2 E\big(\max_{\theta \in \Theta} ||g_{nj}(\theta)||^2 | \Lambda_n \big) \Big)} \sqrt{E \max_{\theta \in \Theta} \Big|\Big| \frac{\partial g_{ni}}{\partial \theta_s}(\theta) \Big|\Big|^2} \\
&\overset{\text{Liapunov}}{\le} n^{-1} \sum_{i,j=1}^{n} \sqrt{E\Big(\lambda_{nij}^2 \sqrt{E\big(\max_{\theta \in \Theta} ||g_{nj}(\theta)||^4 | \Lambda_n \big)} \Big)} \sqrt{E \max_{\theta \in \Theta} \Big|\Big| \frac{\partial g_{ni}}{\partial \theta_s}(\theta) \Big|\Big|^2} \\
&\overset{\text{E}}{\le} \chi_n^{3/4} n^{-1} \sum_{i,j=1}^{n} \sqrt{E \lambda_{nij}^2} \overset{(20)}{\le} \chi_n^{3/4} \Psi_n \overset{\text{E}}{=} O(\Psi_n). \quad (75)
\end{aligned}$$

Repeat the steps of (75) for the remaining RHS summations in (74). ✓

Proof of theorem 4: For $s = 1, \ldots, d$,

$$\begin{aligned}
0 = \frac{\sqrt{n}}{2\Psi_n} \frac{\partial \hat{\Omega}_n}{\partial \theta_s}(\hat{\theta}) &= \sqrt{n} \frac{\partial \bar{g}'_n}{\partial \theta_s}(\hat{\theta}) \Psi_n^{-1} \hat{W}_n(\hat{\theta}) \bar{g}_n(\hat{\theta}) + \frac{\sqrt{n}}{2\Psi_n} \bar{g}'_n(\hat{\theta}) \frac{\partial \hat{W}_n}{\partial \theta_s}(\hat{\theta}) \bar{g}_n(\hat{\theta}) \\
&\overset{\text{L13,L14}}{=} \sqrt{n} \frac{\partial \bar{g}'_n}{\partial \theta_s}(\hat{\theta}) \Psi_n^{-1} \hat{W}_n(\hat{\theta}) \bar{g}_n(\hat{\theta}) + O_p\big(n^{1/2} \Psi_n^{-1} n^{-1/2} \Psi_n^2 n^{-1/2} \big) \overset{\text{I}}{=} \sqrt{n} \frac{\partial \bar{g}'_n}{\partial \theta_s}(\hat{\theta}) \Psi_n^{-1} \hat{W}_n(\hat{\theta}) \bar{g}_n(\hat{\theta}) + o_p(1).
\end{aligned}$$

Hence by the mean value theorem for some $\hat{\theta}^*$ between $\hat{\theta}$ and θ_0,

$$\begin{aligned}
o_p(1) = \sqrt{n} \frac{\partial \bar{g}'_n}{\partial \theta}(\hat{\theta}) \Psi_n^{-1} \hat{W}_n(\hat{\theta}) \bar{g}_n(\hat{\theta}) &= \sqrt{n} \frac{\partial \bar{g}'_n}{\partial \theta}(\hat{\theta}) \Psi_n^{-1} \hat{W}_n(\hat{\theta}) \bar{g}_n(\theta_0) + \sqrt{n} \frac{\partial \bar{g}'_n}{\partial \theta}(\hat{\theta}) \Psi_n^{-1} \hat{W}_n(\hat{\theta}) \frac{\partial g_n}{\partial \theta'}(\hat{\theta}^*)(\hat{\theta} - \theta_0) \\
&\overset{\text{L7,L8,L12,L13}}{=} \sqrt{n} T_0' V_0^{-1} \bar{g}_n(\theta_0) + \sqrt{n} T_0' V_0^{-1} T_0 (\hat{\theta} - \theta_0) + o_p(1),
\end{aligned}$$

which, with assumption J, implies that

$$\sqrt{n}(\hat{\theta}-\theta_0) = (T_0'V_0^{-1}T_0)^{-1}T_0'V_0^{-1}\sqrt{n}\bar{g}_n(\theta_0) + o_p(1) \xrightarrow{\mathcal{D}} N\big(0,(T_0'V_0^{-1}T_0)^{-1}\big),$$

by lemma 7. ✓

B Data Appendix

All variables are yearly and span the 1980-93 period. All monetary variables are in real Canadian dollars, 1993 = 1.00. Mining-industry data, however, are usually reported in U.S. dollars, and people familiar with the industry are accustomed to such numbers. It is therefore helpful to compare the two units. In 1993, a price of U.S. $1.00 per pound was equivalent to CAN $1.35, and costs of U.S. $0.75 were roughly equal to CAN $ 1.00.

CPR: The London Metal Exchange (LME) is the most important exchange for copper trading. Although a copper contract also trades on the Commodity Exchange of New York (COMEX), that market is considerably thinner. For this reason, we use the LME copper price. Yearly prices are averages of the daily grade–A, cash–settlement price published in Nonferrous Metal Data. The Canadian consumer price index, which was obtained from DataStream, and the U.S./Canadian exchange rate, which was obtained from Citibase, were used to convert nominal U.S. to real Canadian cents per pound.

SIGRR: Our measure of volatility is constructed from CPR using equation (29).
These two variables are common to all mines. Other variables vary by mine as well as over time. All mine data are reported on a yearly frequency. The variable COST, however, is available only for years in which the mine operated.

RES: Reserve data, in millions of tonnes of ore, are from the Canadian Mines Handbook.
All other mine data were collected from the Canadian Minerals Yearbook.

CAP: Mine capacity is measured in millions of tonnes of ore per year.

QCU: Metal refined is measured in thousands of tonnes of copper per year.

COST: Average total costs, which include the costs of mining, milling, smelting, refining, shipping, and marketing, are published by Brook Hunt, a consulting firm that specializes in the mining industry. According to industry sources, these costs are the most reliable available and are used extensively by firms in the industry. Fixed and marginal costs, $\hat{FCOST}_i$ and $\hat{MCOST}_i$ for each mine are generated from COST_{it} and QCU_{it} using regression equation (28) applied to active observations.

LAT and *LONG:* Data on location were taken from Infomine, http://www.infomine.com and Canadian Geographic Names, Department of Natural Resources, Canada, http://geonames.nrcan.gc.ca. These data are used to calculate the Euclidean distances between mines that we use to perform the spatial corrections.

The data to produce the outlines of the provinces in Figures 2 and 3 were taken from the Geo Community website http://www.geocomm.com and are in the public domain.

Additional supply and demand variables were used as instruments. These are:

WAGE: A provincial mining wage rate was constructed by dividing the total wage and salary bill for copper/nickel/zinc mines in each province, in thousands of dollars per year, by the number of employees of such mines in that province. The raw wage variables are found in Statistics Canada Catalogue 26-223, table 2.

ENERGY PRICE: A provincial mining energy–price index was constructed as a share–weighted average of the prices of nine classes of fuels that were purchased by copper/nickel/zinc mines in the

province. The raw data consist of two variables for each fuel — the value and quantity of provincial mining–industry purchases. Individual provincial energy prices were obtained by dividing the value by the quantity. These data were then aggregated to form the index. The raw data are found in Statistics Canada Catalogue 26-223 table 6.

INPROD: Canadian industrial–production data, in millions of dollars per year, were obtained from Statistics Canada's computerized data base, Cansim.

C The Normalized Success Index

Suppose that $y_i, i = 1, \ldots, n$ is the observed choice and $\hat{p}_i$ is the predicted probability. Let

$$n_0 = \sum (1 - y_i) \tag{76}$$

and

$$n_1 = \sum y_i \tag{77}$$

be observed counts. Define

$$n_{00} = \sum (1 - y_i)(1 - \hat{p}_i), \tag{78}$$

$$n_{01} = \sum (1 - y_i)\hat{p}_i, \tag{79}$$

$$n_{10} = \sum y_i(1 - \hat{p}_i), \tag{80}$$

and

$$n_{11} = \sum y_i \hat{p}_i. \tag{81}$$

Then

$$\hat{n}_0 = n_{00} + n_{10} \tag{82}$$

and

$$\hat{n}_1 = n_{01} + n_{11} \tag{83}$$

are the predicted counts. The normalized success index is then

$$S = S_0 + S_1 = \left[\frac{n_{00}}{\hat{n}_0} - \frac{n_0}{n}\right] + \left[\frac{n_{11}}{\hat{n}_1} - \frac{n_1}{n}\right]. \tag{84}$$

This is a better measure of goodness of fit than the proportion of successful predictions, which is often used, for the following reason. Suppose that 90% of the observations are zero and 10% are 1. A model that simply predicts zero with probability 0.9 and 1 with probability 0.1 will have a large proportion of successful predictions (approximately 81%). However, it has no predictive power, and its normalized success index is approximately zero.

References Cited

Andrews, D.W.K. (1987) "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica* 55, 1465–1472.

Andrews, D.W.K. (1992) "Generic Uniform Convergence," *Econometric Theory* 8, 241–257.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models,* Kluwer (Dordrecht, The Netherlands).

Arellano, M. and B. Honoré (2001) "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics* V–53, J.J. Heckman and E. Leamer (eds.), pp. 3229–3296.

Bernstein, S. (1927) "Sur l'Extension du Théorème du Calcul des Probabilités aux Sommes de Quantités Dependantes," *Mathematische Annalen* 97, 1–59.

Brennan, M.J. and Schwartz, E.S. (1985) "Evaluating Natural-Resource Assets," *Journal of Business*, 58: 135-157.

Chamberlain, G. (1984) "Panel Data," in *Handbook of Econometrics* volume II, Z. Griliches and M Intriligator, eds, North Holland, Amsterdam.

Cliff, A.D. and J. K. Ord. (1973), "Spatial autocorrelation," Pion, London.

Conley, T.G. (1999) "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics* 92, 1–45.

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

Dixit, A.K. and Pindyck, R.S. (1994) *Investment Under Uncertainty*, Princeton University Press, Princeton, NJ.

Doukhan, P. and S. Louhichi (1999) "A New Weak Dependence Condition and Applications to Moment Inequalities," *Stochastic Processes and their Applications* 84, 312–342.

Hansen, L.P., J. Heaton and A. Yaron (1996) "Finite–Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14, 262–280.

Hensher, D.A. and Johnson, L.W. (1981) *Applied Discrete Choice Modeling,* Wiley, New York.

Honoré , B.E. and E. Kyriazidou (2000) "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 839–874.

Ibragimov, I.A. (1962) "Some Limit Theorems for Stationary Processes," *Theory of Probability and Its Applications* 7, 349–382.

Iglesias, E.M. and G.D.A. Phillips (2005), "Asymptotic Bias of GMM and GEL under Possible Nonstationary Spatial Dependence," Michigan State University mimeo.

Kelejian, H. and I. Prucha (2004) "HAC Estimation in a Spatial Framework," forthcoming in the *Journal of Econometrics*.

Magnac, T. (2004) "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica* 72, 1959–1876.

McDonald, R. and Siegel, D. (1985) "Investment and the Valuation of Firms When There is an Option to Shut Down," *International Economic Review*, 26: 331–349.

Moel, A. and Tufano, P. (2002) "When are Real Options Exercised? An Empirical Study of Mine Closings," *The Review of Financial Studies*, 15: 33–64.

Newey, W.K. and R.J. Smith (2003) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica* 72: 219–255.

Newey, W.K. and K.D. West (1987) "A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55–3:703–708.

Pinkse, J., L. Shen and M.E. Slade (2005), "A Central Limit Theorem for Endogenous Locations and Complex Spatial Interactions," forthcoming in the *Journal of Econometrics.*

Pinkse, J., and Slade, M.E. (1998) "Contracting in Space: An Application of Spatial Statistics to Discrete–Choice Models," *Journal of Econometrics*, 85: 125–154.

Pinkse, J., Slade, M.E. and Brett, C. (2002) "Spatial Price Competition: a Semiparametric Approach," *Econometrica,* 70–3: 1111–1153.

Pollard, D. (1984) *Convergence of Stochastic Processes,* Springer, Berlin.

Pötscher, B. and I. Prucha (1989) "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica* 57, 675–683.

Rosenblatt, M. (1956), "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Science* 42, 43–47.

Slade, M.E. (2001) "Managing Projects Flexibly: An Application of Real-Option Theory to Mining Investments," *Journal of Environmental Economics and Management*, 41: 193-233.

Tourinho, O.A. (1979) "The Valuation of Reserves of Natural Resources: An Option Pricing Approach," Ph.D. dissertation, University of California, Berkeley.

Whittle, P. (1954) *On Stationary Processes in the Plane, Biometrika* 41–3, pp. 434-449.

Table 1: **Summary Statistics**

	All		Active		Inactive	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.
Mine Status (ACTIVE)	0.63	0.48				
Price (CPR)	142.3	32.3	144.9	34.2	137.8	28.2
Volatility of returns (SIGRR)	5.54	2.13	5.48	2.13	5.64	2.14
Reserves remaining (RES)	84.5	166.6	105.3	181.0	49.3	132.2
Capacity (CAP)	5.64	8.70	7.35	10.38	2.74	2.90
Technology (DOPEN)	0.43	0.50	0.49	0.50	0.33	0.47
Marginal cost ($\hat{MCOST}$)	97.6	42.1	91.6	37.0	107.7	48.1
Fixed cost ($\hat{FCOST}$)	31.2	58.2	37.8	58.2	20.1	56.6

Table 2: **Ordinary Probits, GMM, No State Dependence**

#	CPR	SIGRR	RES	CAP	DOPEN	$\hat{MCOST}$	$\hat{FCOST}$	FE	NSI
1	0.0095	-0.115						No	0.03
	(3.2)	(-2.5)							
2	0.0096	-0.118	0.0015					No	0.06
	(3.2)	(-2.5)	(2.4)						
3	0.010	-0.127	-0.0013	0.102				No	0.12
	(3.3)	(-2.7)	(-1.6)	(5.0)					
4	0.0098	-0.120	0.0011		0.237			No	0.06
	(3.3)	(-2.6)	(1.6)		(1.3)				
5	0.010	-0.122	0.0014			-0.0054		No	0.09
	(3.3)	(-2.6)	(2.1)			(-2.8)			
6	0.0097	-0.120	0.0012				0.0028	No	0.07
	(3.2)	(-2.6)	(2.0)				(1.8)		
7	0.016	-0.201						Yes	0.47
	(3.4)	(-3.2)							
8	0.016	-0.201	0.0005					Yes	0.48
	(3.3)	(-3.1)	(0.1)						
9	0.016	-0.197	-0.0021	0.109				Yes	0.49
	(3.4)	(-3.2)	(-0.3)	(1.0)					

Dependent variable: ACTIVE.
t statistics in parentheses.
FE denotes specifications with mine fixed effects.
Fixed effects included in latent–variable equation.
NSI denotes normalized success index.

Table 3: **Ordinary Probits, GMM, State–Dependent Estimates**

#	CPR	SIGRR1	RES	CAP1	SIGRR0	CAP0	FE	NSI
1	0.019	-0.085			-0.685		No	0.67
	(2.8)	(-1.1)			(-6.0)			
2	0.019	-0.093	0.0012		-0.689		No	0.67
	(2.8)	(-1.2)	(1.4)		(-6.0)			
3	0.019	-0.093	-0.0007	0.075	-0.670	0.070	No	0.68
	(2.8)	(-1.1)	(-0.6)	(1.9)	(-5.3)	(1.0)		
4	0.022	-0.162			-0.674		Yes	0.74
	(3.3)	(-1.8)			(-5.3)			
5	0.023	-0.169	0.0025		-0.686		Yes	0.74
	(3.4)	(-1.9)	(0.9)		(-5.3)			
6	0.024	-0.175	-0.0010	0.175	-0.718	0.207	Yes	0.74
	(3.4)	(-1.9)	(-0.2)	(1.0)	(-4.9)	(1.2)		

Dependent variable: ACTIVE.
t statistics in parentheses.
FE denotes specifications with mine fixed effects.
Fixed effects included in latent–variable equation.
Variables ending in 1 (0) correspond to observations with the mine previously open (closed).
NSI denotes normalized success index.

Table 4: **Heteroskedasticity Corrected Models: No State Dependence**

#	CPR	SIGRR	RES	CAP	FE	NSI
1	0.0075	-0.084			No	0.02
	(1.7)	(-1.4)				
2	0.0087	-0.112	0.0014		No	0.05
	(1.9)	(-1.8)	(2.6)			
3	0.0088	-0.099	-0.0015	0.117	No	0.12
	(1.9)	(-1.5)	(-1.7)	(5.0)		
4	0.0082	-0.182			Yes	
	(0.6)	(-0.9)				
5	0.014	-0.238	0.0017		Yes	
	(1.0)	(-1.1)	(1.1)			
6	0.0037	-0.089	-0.010	0.232	Yes	
	(0.5)	(-0.5)	(-0.9)	(1.1)		

Dependent variable: ACTIVE.
t statistics in parentheses.
FE denotes specifications with mine fixed effects.
NSI denotes normalized success index.

Table 5: **Heteroskedasticity Corrected Models: State Dependence**

#	CPR	SIGRR1	RES	CAP1	SIGRR0	CAP0	FE	NSI
1	0.039	-0.342			-2.53		No	0.74
	(1.3)	(-1.5)			(-1.3)			
2	0.058	-0.468	0.0013		-3.55		No	0.74
	(1.6)	(-1.8)	(1.6)		(-1.5)			
3	0.057	-0.329	0.0053	0.076	-3.88	0.062	No	0.79
	(1.3)	(-1.2)	(0.8)	(0.9)	(-1.3)	(0.1)		
4	0.075	-0.857			-4.73		Yes	
	(0.5)	(-0.7)			(-0.5)			
5	0.076	-0.854	0.0007		-4.89		Yes	
	(0.7)	(-0.8)	(0.7)		(-0.7)			
6	0.140	-1.26	-0.063	2.94	-7.48	0.941	Yes	
	(0.2)	(-0.2)	(-0.3)	(0.3)	(-0.2)	(0.4)		

Dependent variable: ACTIVE.
t statistics in parentheses.
FE denotes specifications with mine fixed effects.
Variables ending in 1 (0) correspond to observations with the mine previously open (closed).
NSI denotes normalized success index.

Table 6: **Spatial Models, No State Dependence**

#	CPR	SIGRR	RES	CAP	FE	NSI
1	0.0033	-0.042			No	0.01
	(0.8)	(-0.8)				
2	0.0058	-0.071	0.0013		No	0.04
	(1.4)	(-1.3)	(2.0)			
3	0.0028	-0.018	-0.0018	0.115	No	0.10
	(0.6)	(-0.3)	(-1.5)	(4.1)		
4	-0.0007	-0.579			Yes	
	(-0.06)	(-0.7)				
5	0.0009	-0.671	0.0014		Yes	
	(0.06)	(-0.8)	(1.0)			
6	0.0051	-0.123	-0.016	0.340	Yes	
	(0.7)	(-0.8)	(-1.2)	(1.3)		

Dependent variable: ACTIVE.
t statistics in parentheses.
F.E. denotes specifications with mine fixed effects.
NSI denotes normalized success index.

Table 7: **Spatial Models: State Dependence**

#	CPR	SIGRR1	RES	CAP1	SIGRR0	CAP0	FE	NSI
1	0.045	-0.399			-2.81		No	0.74
	(1.5)	(-1.8)			(-1.5)			
2	0.056	-0.472	0.0012		-3.45		No	0.74
	(1.7)	(-2.0)	(2.0)		(-1.6)			
3	0.057	-0.338	0.0024	0.101	-4.02	0.163	No	0.79
	(1.5)	(-1.2)	(0.4)	(1.3)	(-1.6)	(0.4)		
4	0.051	-0.523			-3.48		Yes	
	(1.0)	(-1.2)			(-1.0)			
5	0.058	-0.620	0.0005		-4.01		Yes	
	(1.0)	(-1.2)	(0.5)		(-1.0)			
6	0.053	-0.615	-0.031	1.98	-2.57	0.614	Yes	
	(1.9)	(-1.9)	(-1.6)	(1.5)	(-1.3)	(1.3)		

Dependent variable: ACTIVE.
t statistics in parentheses.
FE denotes specifications with mine fixed effects.
Variables ending in 1 (0) correspond to observations with the mine previously open (closed).
NSI denotes normalized success index.

Table 8: **Predicted Transition Probabilities**

Previous State		
	at Mean	
	OPEN	CLOSED
OPEN	0.91	0.09
CLOSED	0.07	0.93
	High Volatility	
	OPEN	CLOSED
OPEN	0.88	0.12
CLOSED	0.004	0.994

Top half: All explanatory variables at mean.
Bottom half: High volatility equals mean plus one standard deviation
All other explanatory variables at mean

Figure 2: Mine locations in Quebec

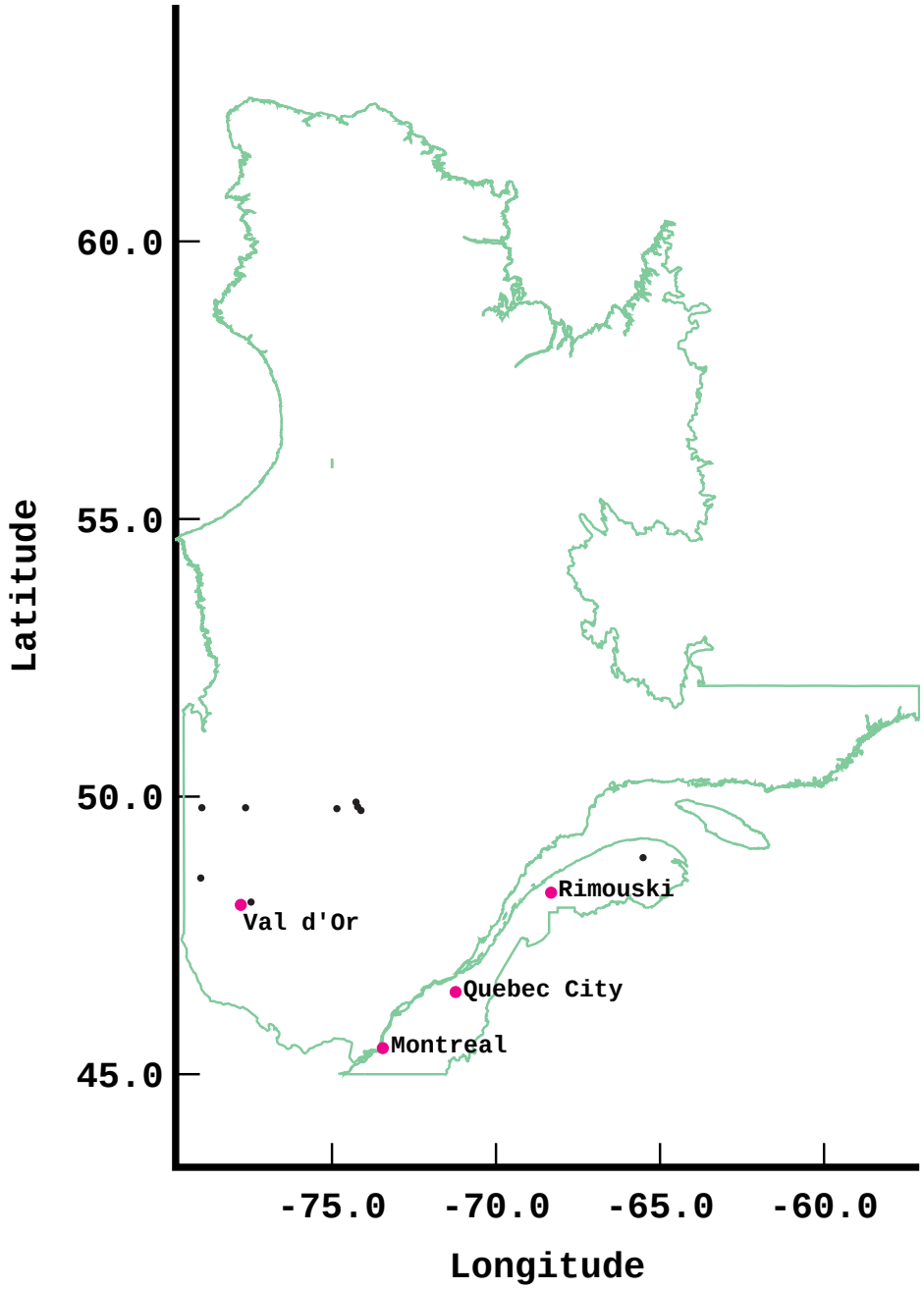

Figure 3: Mine locations in British Columbia