Time Invariant Regressor in Nonlinear Panel Model with Fixed Effects\textsuperscript{1}

Jinyong Hahn
UCLA
February 26, 2003

\textsuperscript{1}I am grateful to Dan Ackerberg and Jerry Hausman for helpful comments.
Abstract

This paper generalizes Hausman and Taylor’s (1981) intuition, and develop a method of dealing with time-invariant regressor in nonlinear panel model with fixed effects. The method requires large number of observations per individual ($T$) as well as a large number of individuals ($n$).
1 Introduction

Panel data allow the possibility of controlling for unobserved individual specific effects, which may be correlated with observed explanatory variables. In linear models, such “fixed effects” are usually eliminated by differencing, which yields a model free of such incidental parameter. An unintended consequence of differencing is that it also eliminates time-invariant regressor, which renders the coefficient of the time-invariant regressor unidentified. Hausman and Taylor (1981) used an instrumental variables approach to overcome such problem.

In this paper, I generalize their intuition, and develop a method of dealing with time-invariant regressor in the nonlinear framework. This method requires large number of observations per individual (T), so its applicability is limited to the case where T is large. Because the instrumental variables estimation requires a large number of individuals (n), I adopt a asymptotic framework where both n and T grow to infinity at the same rate. This result is made possible by recent technical progress of panel analysis under such alternative asymptotics. See, e.g., Arellano (2000), Hahn and Kuersteiner (2002, 2003), Hahn and Newey (2002), and Woutersen (2002).

2 Preliminaries

Suppose that we are given a set of moment restrictions

\[ E [g (y_{it}, \gamma_{i0} + w_i' \delta_0 + x_{it}' \theta_0)] = 0, \quad i = 1, \ldots, n; t = 1, \ldots, T \]

for some vector-valued function g, where \( y_{it}, x_{it}, \) and \( w_i \) denote the dependent variable in the tth period, time-varying regressor in the tth period, and time-invariant regressor. Unobserved individual specific effects are summarized by the scalar variable \( \gamma_i \). Our primary focus is to estimate the coefficient \( \delta_0 \) of the time-invariant regressor \( w_i \) when \( \gamma_i \) is possibly correlated with \( w_i \) and \( x_{it} \).

We now discuss how the parameters can be consistently estimated. If both n and T grow to infinity at the same rate, \( \theta_0 \) can be \( \sqrt{nT} \)-consistently estimated. Letting

\[ \alpha_{i0} \equiv \gamma_{i0} + w_i' \delta_0 \]

we can rewrite the model as

\[ E [g (y_{it}, \alpha_{i0} + x_{it}' \theta_0)] = 0, \]

to which we can apply variants of recently developed methods discussed in Arellano (2000), Hahn and Kuersteiner (2002, 2003), Hahn and Newey (2002), and Woutersen (2002).

Therefore, the conceptual challenge is to develop a consistent estimator of \( \delta_0 \). For this purpose, we assume that the data are i.i.d. over i:

**Condition 1** \((\{y_{i1}, y_{i2}, \ldots\}, \{x_{i1}, x_{i2}, \ldots\}, z_i, w_i, \gamma_{i0}\) is i.i.d. over i.

Suppose for a moment that we observe \( \alpha_{i0} \). Also suppose that we observe an additional variable \( z_i \) with \( \dim (z_i) = \dim (w_i) \)

\[ ^1 \text{It is easy to generalize the discussion to the over-identified case where dim (z_i) > dim (w_i). Because the primary purpose of this paper is identification and consistent estimation of } \delta_0, \text{ I focus on the exactly identified case.} \]
Condition 2 (i) $E[z_i\gamma_{i0}] = 0$; (ii) $E[z_iw_i]$ is nonsingular

It is clear that we can estimate $\delta_0$ by $\hat{\delta} \equiv (\sum_{i=1}^{n} z_iw_i)^{-1} (\sum_{i=1}^{n} z_i\alpha_{i0})$. It is easy to see that this estimator is consistent for $\delta_0$ as $n \to \infty$. Hausman and Taylor’s (1981) intuition was that $\hat{\delta}$ would remain consistent even if we replace $\alpha_{i0}$ by an unbiased estimate. In the nonlinear context, it seems difficult to come up with such an unbiased estimator for $\alpha_{i0}$. Therefore, Hausman and Taylor’s (1981) method cannot be directly applied.

The basic intuition in this paper is that, when both $n$ and $T$ grow to infinity at the same rate, we can come up with a $\sqrt{T}$-consistent estimator for $\alpha_{i0}$, say $\hat{\alpha}_i$. Because the estimation error becomes very smaller as the sample size increases, the IV estimator

$$\hat{\delta} \equiv (\sum_{i=1}^{n} z_iw_i)^{-1} (\sum_{i=1}^{n} z_i\hat{\alpha}_i)$$

will be consistent for $\delta_0$ in general.

3 Consistent Estimation of $\alpha_{i0}$

Let

$$g(X_{it}; \theta, \alpha_i) \equiv \begin{bmatrix} u(X_{it}; \theta, \alpha_i) \\ v(X_{it}; \theta, \alpha_i) \end{bmatrix}$$

where $X_{it} \equiv (y_{i1}, \ldots, y_{iT}, x_{i1}, \ldots, x_{iT})$, and where $\dim(\theta) = \dim(u) = p$ and $\dim(\alpha) = \dim(v) = 1$. We assume that

$$E[g(X_{it}; \theta_0, \alpha_{i0})] = 0,$$

and consider the estimator that solves

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{T} u\left(X_{it}; \hat{\theta}, \hat{\alpha}_i \right)$$

$$0 = \sum_{t=1}^{T} v\left(X_{it}; \hat{\theta}, \hat{\alpha}_i \right)$$

This indicates that we are separating $g$ into two components. The first component $u$ is used throughout the sample for estimation of $\theta_0$. The second component $v$ is used only for the $i$th individual to estimate $\alpha_{i0}$. This separation was adopted because I wanted to exploit some recent technical development in Hahn and Kuersteiner (2003) or Hahn and Newey (2002). I do not expect this separation to be constraining in practice. For example, in the case of linear models

$$y_{it} = \alpha_{i0} + x_{it}'\theta_0 + \varepsilon_{it}$$

(2)

with $x_{it}$ strictly exogenous, we may take

$$u(X_{it}; \theta, \alpha_i) = x_{it} \cdot (y_{it} - \alpha_i - x_{it}'\theta)$$

$$v(X_{it}; \theta, \alpha_i) = y_{it} - \alpha_i - x_{it}'\theta$$
where \( \bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it} \) and \( \bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it} \). Note that this will result in the usual fixed effects estimator of \( \theta_0 \).

We assume following regularity conditions:

**Condition 3** (i) Given time-invariant variables \((\alpha_{i0}, z_i, w_i)\), \((y_{it}, x_{it})\) is i.i.d. over \( t \); (ii) for every \( i \), \( G_i(\theta, \alpha_{i0}) = 0 \); (iii) for each \( n > 0 \), \( \inf \inf G_i(\theta, \alpha_{i0}) > 0 \), where \( \bar{G}_i(\theta, \alpha_{i0}) = T^{-1} \sum_{t=1}^{T} g \left( X_{it}; \theta, \alpha_{i0} \right) \), \( G_i(\theta, \alpha_{i0}) \equiv E \left[ g \left( X_{it}; \theta, \alpha_{i0} \right) \right] \).

**Condition 5** (i) The function \( g(\cdot; \theta, \alpha) \) is continuous in \((\theta, \alpha) \in Y\); (ii) The parameter space \( Y \) is compact; (iii) There exists a function \( M(X_{it}) \) such that

\[
\left| \frac{\partial m_1 + m_2 g(X_{it}; \theta, \alpha_{i0})}{\partial \alpha M_1 \partial \alpha M_2} \right| \leq M(X_{it}) \quad 0 \leq m_1 + m_2 \leq 1, \ldots, 6
\]

and \( \sup \left[ E \left[ M(X_{it})^Q \right] \right] < \infty \) for some \( Q > 64 \).

**Condition 6** (i) \( \min \left[ E \left[ v(X_{it}; \theta, \alpha_{i0})^2 \right] > 0 \right] \); (ii) \( \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{I}_t = 0 \), where \( U(X_{it}; \theta, \alpha_{i0}) \equiv u(X_{it}; \theta, \alpha_{i0}) - \rho_{i0} v(X_{it}; \theta, \alpha_{i0}), \rho_{i0} \equiv E \left[ \frac{\partial u(X_{it}; \theta, \alpha_{i0})}{\partial \alpha_{i0}} \right] \). \( \mathcal{I}_t = -E \left[ \frac{\partial u(X_{it}; \theta, \alpha_{i0})}{\partial \theta} \right] \).

We can show that \( \hat{\alpha}_i \) are uniformly consistent for \( \alpha_{i0} \).

**Theorem 1** Under Conditions 3, 4, 5, and 6, we have

\[
\sqrt{T} (\hat{\alpha}_i - \alpha_{i0}) = \lambda_i + \kappa_i.
\]

where

\[
\lambda_i = - \left( E \left[ \frac{\partial \nu_{it}}{\partial \alpha} \right] \right)^{-1} \left( T^{-1/2} \sum_{t=1}^{T} v_{it} \right) = O_p(1)
\]

and \( \operatorname{Pr} \left[ \max \left| \kappa_i \right| \right] = o(1) \). Here, \( v_{it} \equiv v_{it} (X_{it}, \theta_0, \alpha_{i0}) \).

**Proof.** See Appendix C.

---

4 IV Estimation of \( \delta_0 \)

Recall that, if \( \alpha_{i0} \) is known, we could have estimate \( \delta_0 \) by

\[
\hat{\delta} \equiv \left( \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left( \sum_{i=1}^{n} z_i \alpha_{i0} \right).
\]

Because it is not known, we can replace \( \alpha_{i0} \) by \( \hat{\alpha}_i \), and consider

\[
\hat{\delta} \equiv \left( \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left( \sum_{i=1}^{n} z_i \hat{\alpha}_i \right).
\]
Theorem 1 implies that we have
\[
\sqrt{n} (\hat{\delta} - \delta_0) = \left( n^{-1} \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left( n^{-1/2} \sum_{i=1}^{n} z_i \gamma_{i0} + n^{-1/2} T^{-1/2} \sum_{i=1}^{n} z_i \lambda_i \right) + o_p(1)
\]

Note that, if \( z_i \lambda_i = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_i v_{it} \) does not have a zero expectation, it would complicate asymptotic analysis. We therefore assume that the instrument \( z_i \) satisfies \( E \left[ z_i \cdot v_{it} \right] = 0 \).

**Condition 7** \( E \left[ v_{it} | z_i \right] = 0 \)

Condition 7 guarantees that the error in estimation of \( \alpha_{i0} \) does not complicate the inference regarding \( \hat{\delta} \). Such condition was adopted by Hausman and Taylor (1981). We have \( v_{it} = \varepsilon_{it} \) in the linear model (2), and Hausman and Taylor’s instrument \( z_i \) for such linear model is required to satisfy \( E \left[ z_i \cdot \varepsilon_{it} \right] = 0 \).

It turns out that the substitution of \( \hat{\alpha}_i \) for \( \alpha_{i0} \) in (3) does not affect the asymptotic distribution of the resultant estimator (4). It is because \( n^{-1/2} T^{-1/2} \sum_{i=1}^{n} z_i \lambda_i = o_p(1) \) under Conditions 3, and 7. We therefore obtain
\[
\sqrt{n} (\hat{\delta} - \delta_0) = \left( n^{-1} \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left( n^{-1/2} \sum_{i=1}^{n} z_i \gamma_{i0} \right) + o_p(1)
\]

Note that
\[
\sqrt{n} (\hat{\delta} - \delta_0) = \left( n^{-1} \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left( n^{-1/2} \sum_{i=1}^{n} z_i \gamma_{i0} \right) + o_p(1)
\]
as well. This is an intuitive result. Note that \( \hat{\delta} \) is an IV estimator of \( \hat{\alpha}_i \) on \( w_i \). Because \( \hat{\alpha}_i \) is a proxy for \( \alpha_{i0} \), and because the “measurement error” disappears as \( T \to \infty \), we should expect that the asymptotic distribution of \( \sqrt{n} (\hat{\delta} - \delta_0) \) should be identical to that of \( \sqrt{n} (\tilde{\delta} - \delta_0) \) under the large \( T \) asymptotics.

As a consequence, we obtain

**Theorem 2** Assume Conditions 1, 2, 3, 4, 5, 6, and 7. Further assume that \( E \left[ |z_i w_i'| \right] < \infty \) and \( E \left[ |\gamma_{i0}^2 z_i z_i'| \right] < \infty \). We then have
\[
\sqrt{n} (\hat{\delta} - \delta_0) \to N \left( 0, (E \left[ z_i w_i' \right])^{-1} E \left[ \gamma_{i0}^2 z_i z_i' \right] (E \left[ w_i z_i' \right])^{-1} \right).
\]

5 Summary and Future Work

In this paper, I generalized Hausman and Taylor’s (1981) result to nonlinear panel models with fixed effects. The result is expected to have an implication for some nonlinear models of social interactions. This is because the model (1) could be understood as a model with some particular kinds of social effects. Following the terminology introduced by Manski (1993), let \( y_{it} \) denote the outcome of the \( t \)th individual of the \( i \)th group. Also, let \( x_{it} \) denote the individual level characteristics that operate at the individual level only. If \( w_i \) is equal to the group average \( E_i[y_{it}] \) of \( y_{it} \), then the model contains endogenous social effects. If \( w_i \) is equal to the group average \( E_i[x_{it}] \) of \( x_{it} \), then the model contains exogenous (contextual) social effects. Finally \( \gamma_{i0} \), which is not observed by the econometrician, captures the presence of correlated
group effects. Therefore, the result in this paper may be applicable to some particular class of models of social interactions.²

²Identification of $\delta_0$ with various kinds of social effects is expected to require some more functional restriction, such as the exclusion restriction considered by Graham and Hahn (2003), or selection effects as considered by Brock and Durlauf (2001). Nonlinear models of social interaction are expected to be plagued by multiple equilibria in general. Therefore, it is important that the model (1) is not understood as the first order condition of the log likelihood. After all, textbook style likelihood may not exit. The model (1) should be understood as a set of moment restrictions that are valid regardless of the potential multiplicity of equilibria. With multiple equilibria, the i.i.d. assumption in Condition 1 may sound puzzling. Such puzzle would disappear if the equilibrium selection is considered as a fixed effect, although it does not appear in the model (1).
Appendix

A Consistency

Lemma 1 Assume that \(W_t\) are iid with \(E[W_t] = 0\) and \(E[W_t^{2k}] < \infty\). Then,

\[
E \left[ \left( \sum_{t=1}^{T} W_t \right)^{2k} \right] = C(k)T^k + o(T^k)
\]

for some constant \(C(k)\).

Proof. By adopting an argument in the proof of Lemma 5.1 in Lahiri (1992), we have

\[
E \left[ \left( \sum_{t=1}^{T} W_t \right)^{2k} \right] = 2^k \sum_{\alpha} C(\alpha_1, ..., \alpha_j) \sum_{\alpha} E \left[ \prod_{s=1}^{j} W_{t_s}^{\alpha_s} \right],
\]

where for each fixed \(j \in \{1, ..., 2k\}\), \(\sum_{\alpha}\) extends over all \(j\)-tuples of positive integers \((\alpha_1, ..., \alpha_j)\) such that \(\alpha_1 + ... + \alpha_j = 2k\) and \(\sum_{\alpha}\) extends over all ordered \(j\)-tuples \((t_1, ..., t_j)\) of integers such that \(1 \leq t_j \leq T\). Also, \(C(\alpha_1, ..., \alpha_j)\) stands for a bounded constant. Note, that if \(j > k\) then at least one of the indices \(\alpha_j = 1\). By independence and the fact that \(E[W_t] = 0\) it follows that \(E \left[ \prod_{s=1}^{j} W_{t_s}^{\alpha_s} \right] = 0\) whenever \(j > k\).

This shows that \(E \left[ \left( \sum_{t=1}^{T} W_t \right)^{2k} \right] \leq C(k)T^k \cdot E \left[ W_t^{2k} \right]\) for some constant \(C(k)\). □

Lemma 2 Suppose that, for each \(i\), \(\{\xi_{it}, t = 1, 2, \ldots\}\) is a sequence of zero mean i.i.d. random variables. We assume that \(\{\xi_{it}, t = 1, 2, 3\}\) are independent across \(i\). We also assume that \(\max_i E \left| \xi_{it} \right|^{16} < \infty\). Finally, we assume that \(n = O(T)\). We then have

\[
\max_i \Pr \left[ \left| \frac{1}{T} \sum_{t=1}^{T} \xi_{it} \right| > \eta \right] = o(T^{-2})
\]

for every \(\eta > 0\).

Proof. Using Lemma 1, we obtain

\[
E \left[ \left| \sum_{t=1}^{T} \xi_{it} \right|^{16} \right] \leq CT^{8} \cdot E \left[ \xi_{it}^{16} \right],
\]

where \(C > 0\) is a constant. Therefore, we have

\[
T^{2} \Pr \left[ \left| \frac{1}{T} \sum_{t=1}^{T} \xi_{it} \right| > \eta \right] \leq T^{2} \frac{CT^{8}}{T^{16} \eta^{16}} E \left[ \xi_{it}^{16} \right]
\]

or

\[
\max_i T^{2} \Pr \left[ \left| \frac{1}{T} \sum_{t=1}^{T} \xi_{it} \right| > \eta \right] \leq \frac{C}{T^{8} \eta^{16}} \max_i E \left[ \xi_{it}^{16} \right] = o(1).
\]

□
Lemma 3 Suppose that Conditions 4 and 5 hold. We then have for all $\eta > 0$ that

$$\Pr \left[ \max_{1 \leq i \leq n} \sup_{(\theta, \alpha)} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \geq \eta \right] = o(T^{-1})$$

Proof. Let $\eta > 0$ be given. We note that

$$\Pr \left[ \max_{1 \leq i \leq n} \sup_{(\theta, \alpha)} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \geq \eta \right] \leq \sum_{i=1}^{n} \Pr \left[ \sup_{(\theta, \alpha)} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \geq \eta \right]$$

(6)

Let $\varepsilon > 0$ be chosen such that $2\varepsilon \max_{i} E [M (X_{it})] < \frac{\eta}{3}$. Divide $\Upsilon$ into subsets $\Upsilon_{1}, \Upsilon_{2}, \ldots, \Upsilon_{M(\varepsilon)}$ such that $| (\theta, \alpha) - (\theta', \alpha') | < \varepsilon$ whenever $(\theta, \alpha)$ and $(\theta', \alpha')$ are in the same subset. Let $(\theta_{j}, \alpha_{j})$ denote some point in $\Upsilon_{j}$ for each $j$. Then,

$$\sup_{(\theta, \alpha)} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| = \max_{j} \sup_{\Upsilon_{j}} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right|,$$

and therefore

$$\Pr \left[ \sup_{(\theta, \alpha)} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \geq \eta \right] \leq \sum_{j=1}^{M(\varepsilon)} \Pr \left[ \sup_{\Upsilon_{j}} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \geq \eta \right]$$

(7)

For $(\theta, \alpha) \in \Upsilon_{j}$, we have

$$\left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \leq \left| \widehat{G}_{(i)} (\theta_{j}, \alpha_{j}) - G_{(i)} (\theta_{j}, \alpha_{j}) \right| + \frac{\varepsilon}{T} \left| \sum_{t=1}^{T} (M (X_{it}) - E [M (X_{it})]) \right| + 2\varepsilon E [M (X_{it})]$$

Then,

$$\Pr \left[ \sup_{\Upsilon_{j}} \left| \widehat{G}_{(i)} (\theta, \alpha) - G_{(i)} (\theta, \alpha) \right| \geq \eta \right] \leq \Pr \left[ \left| \widehat{G}_{(i)} (\theta_{j}, \alpha_{j}) - G_{(i)} (\theta_{j}, \alpha_{j}) \right| \geq \frac{\eta}{3} \right] + \Pr \left[ \frac{1}{T} \left| \sum_{t=1}^{T} (M (X_{it}) - E [M (X_{it})]) \right| \geq \frac{\eta}{3\varepsilon} \right]$$

(8)

by Lemma 2. Combining (6), (7), (8), and $n = O(T)$, we obtain the desired conclusion. ■

Theorem 3 Under Conditions 4, 5, and 3, $\Pr \left[ \left| \theta - \theta_{0} \right| \geq \eta \right] = o(T^{-1})$ for every $\eta > 0$.

Proof. It is useful to note $(\widehat{\theta}, \widehat{\alpha}_{1}, \ldots, \widehat{\alpha}_{n})$ solves $\widehat{G}_{(i)} (\theta, \alpha_{i}) = 0$ for every $i$. In other words, $(\widehat{\theta}, \widehat{\alpha}_{1}, \ldots, \widehat{\alpha}_{n})$ solves $\min_{\theta} - \sum_{i=1}^{n} \left| \widehat{G}_{(i)} (\theta, \alpha_{i}) \right|$. Let $\eta$ be given, and let $\varepsilon \equiv \inf_{\theta_{0}} \inf_{\left| (\theta, \alpha) - (\theta_{0}, \alpha_{0}) \right| > \eta} \left| G_{(i)} (\theta, \alpha) \right|$. Note that $\varepsilon > 0$. With probability equal to $1 - o(T^{-1})$, we have

$$\min_{\left| \theta - \theta_{0} \right| > \eta, \alpha_{1}, \ldots, \alpha_{n}} \left| \sum_{i=1}^{n} \left| \widehat{G}_{(i)} (\theta, \alpha_{i}) \right| \right| \geq \min_{\left| (\theta, \alpha) - (\theta_{0}, \alpha_{0}) \right| > \eta} \left| \sum_{i=1}^{n} \left| \widehat{G}_{(i)} (\theta, \alpha_{i}) \right| \right| > \min_{\left| (\theta, \alpha) - (\theta_{0}, \alpha_{0}) \right| > \eta} \left| \sum_{i=1}^{n} \left| G_{(i)} (\theta, \alpha_{i}) \right| \right| - \frac{1}{2}\varepsilon$$

$$> \frac{1}{2}\varepsilon$$

$$> 0$$

$$= \sum_{i=1}^{n} \left| \widehat{G}_{(i)} (\theta, \alpha_{i}) \right|,$$
where the second and fourth inequalities are based on Lemma 3, and the last equality follows from the definition of the estimators \( (\hat{\theta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_n) \). We can therefore conclude that \( \Pr \left[ |\hat{\theta} - \theta_0| \geq \eta \right] = o(T^{-1}) \).

\[ \text{Theorem 4} \quad \text{Under Conditions 4, 5, and 3,} \quad \Pr \left[ \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| \geq \eta \right] = o(T^{-1}) \quad \text{for every} \ \eta > 0. \]

\[ \text{Proof.} \quad \text{We first prove that} \]
\[ T \Pr \left[ \max_{1 \leq i \leq n} \sup_{\alpha} \left| \hat{G}_i(\hat{\theta}, \alpha) - G_i(\theta_0, \alpha) \right| \geq \eta \right] = o(1) \]
\[ \text{for every} \ \eta > 0. \ \text{Note that} \]
\[ \max_{1 \leq i \leq n} \sup_{\alpha} \left| \hat{G}_i(\hat{\theta}, \alpha) - G_i(\theta_0, \alpha) \right| \leq \max_{1 \leq i \leq n} \sup_{\alpha} \left| \hat{G}_i(\hat{\theta}, \alpha) - G_i(\hat{\theta}, \alpha) \right| + \max_{1 \leq i \leq n} \sup_{\alpha} \left| G_i(\hat{\theta}, \alpha) - G_i(\theta_0, \alpha) \right| \]
\[ \leq \max_{1 \leq i \leq n} \sup_{(\theta, \alpha)} \left| \hat{G}_i(\theta, \alpha) - G_i(\theta, \alpha) \right| + \max_{1 \leq i \leq n} E\{M(X_0)\} \cdot |\hat{\theta} - \theta_0|. \]

Therefore,
\[ T \Pr \left[ \max_{1 \leq i \leq n} \sup_{\alpha} \left| \hat{G}_i(\hat{\theta}, \alpha) - G_i(\theta_0, \alpha) \right| \geq \eta \right] \leq T \Pr \left[ \max_{1 \leq i \leq n} \sup_{(\theta, \alpha)} \left| \hat{G}_i(\theta, \alpha) - G_i(\theta, \alpha) \right| \geq \eta \right] \]
\[ + T \Pr \left[ |\hat{\theta} - \theta_0| \geq \frac{\eta}{2(1 + \max_{1 \leq i \leq n} E\{M(X_0)\})} \right] = o(1) \]

by Lemma 3 and Theorem 3.

We now get back to the proof of Theorem 4. It suffices to prove that

\[ T \Pr \left[ \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| \geq \eta \right] = o(1) \]

for every \( \eta > 0 \). Let \( \eta \) be given, and let \( \varepsilon \equiv \inf_i \inf_{|\alpha_i - \alpha_0| > \eta} |G_i(\theta_0, \alpha_i)| > 0 \). Condition on the event \( \{ \max_{1 \leq i \leq n} \sup_{\alpha} \left| \hat{G}_i(\hat{\theta}, \alpha) - G_i(\theta_0, \alpha) \right| \leq \frac{1}{3} \varepsilon \} \), which has a probability equal to \( 1 - o(T^{-1}) \) by (9). We then have

\[ \min_{|\alpha_i - \alpha_0| > \eta} \left| \hat{G}_i(\hat{\theta}, \alpha_i) \right| > \min_{|\alpha_i - \alpha_0| > \eta} \left| G_i(\theta_0, \alpha_i) \right| - \frac{1}{2} \varepsilon > \frac{1}{2} \varepsilon > 0 = \left| \hat{G}_i(\hat{\theta}, \hat{\alpha}_i) \right|, \]

and therefore, \( |\hat{\alpha}_i - \alpha_{i0}| \leq \eta \) for every \( i \).
B Expansion

Letting

\[
U(X_{it}; \theta; \alpha_i) = u(X_{it}; \theta; \alpha_i) - \rho_{i0} v(X_{it}; \theta; \alpha_i)
\]

\[
\rho_{i0} = E \left( \frac{\partial u(X_{it}; \theta; \alpha_i)}{\partial \alpha_i'} \right) \left( E \left( \frac{\partial v(X_{it}; \theta; \alpha_i)}{\partial \alpha_i'} \right) \right)^{-1}
\]

\[
0 = \sum_{t=1}^{T} v(X_{it}; \theta; \hat{\alpha}_i(\theta))
\]

\[
V(X_{it}; \theta; \alpha_i) = v(X_{it}; \theta; \alpha_i)
\]

we can recognize that \( \hat{\theta} \) is a solution to

\[
0 = \sum_{i=1}^{n} \sum_{t=1}^{T} U(X_{it}; \hat{\theta}, \hat{\alpha}_i(\hat{\theta})).
\]

Let \( F \equiv (F_1, \ldots, F_n) \) denote the collection of distribution functions. Let \( \hat{F} \equiv (\hat{F}_1, \ldots, \hat{F}_n) \), where \( \hat{F}_i \) denotes the empirical distribution function for the stratum \( i \). Define \( F(\varepsilon) \equiv F + \varepsilon \sqrt{T} \left( \hat{F} - F \right) \) for \( \varepsilon \in [0, T^{-1/2}] \). For each fixed \( \theta \) and \( \varepsilon \), let \( \alpha_i(\theta, F_i(\varepsilon)) \) be the solution to the estimating equation

\[
0 = \int V_i(\theta, \alpha_i(\theta, F_i(\varepsilon))) \, dF_i(\varepsilon),
\]

and let \( \theta(F(\varepsilon)) \) be the solution to the estimating equation

\[
0 = \sum_{i=1}^{n} \int U_i(X_{it}; \theta(F(\varepsilon)), \alpha_i(\theta(F_i(\varepsilon)), F_i(\varepsilon))) \, dF_i(\varepsilon).
\]

By Taylor series expansion, we have

\[
\theta\left(\hat{F}\right) - \theta(F) = \frac{1}{\sqrt{T}} \theta''(0) + \frac{1}{2} \left( \frac{1}{\sqrt{T}} \right)^2 \theta''(0) + \frac{1}{6} \left( \frac{1}{\sqrt{T}} \right)^3 \theta'''(\bar{\varepsilon}),
\]

where \( \theta''(\varepsilon) \equiv d\theta(F(\varepsilon))/d\varepsilon, \theta'''(\varepsilon) \equiv d^2\theta(F(\varepsilon))/d\varepsilon^2, \ldots \), and \( \bar{\varepsilon} \) is somewhere in between 0 and \( T^{-1/2} \).

We therefore have

\[
\sqrt{nT} \left( \theta\left(\hat{F}\right) - \theta(F) \right) = \sqrt{nT} \frac{1}{\sqrt{T}} \theta''(0) + \sqrt{nT} \frac{1}{2} \left( \frac{1}{\sqrt{T}} \right)^2 \theta''(0) + \frac{1}{6} \sqrt{nT} \frac{1}{\sqrt{T}} \theta'''(\bar{\varepsilon}). \tag{10}
\]

The last term in (10) can be shown to be \( o_p(1) \) by the same method as in Hahn and Newey (2002).

Let

\[
h_i(\cdot, \varepsilon) \equiv U_i(\cdot; \theta(F(\varepsilon)), \alpha_i(\theta(F(\varepsilon)), F_i(\varepsilon))) \tag{11}
\]

The first order condition may be written as

\[
0 = \frac{1}{n} \sum_{i=1}^{n} \int h_i(\cdot, \varepsilon) \, dF_i(\varepsilon) \tag{12}
\]
Differentiating repeatedly with respect to \( \epsilon \), we obtain
\[
0 = \frac{1}{n} \sum_{i=1}^{n} \int \frac{dh_i(\cdot, \epsilon)}{d\epsilon} dF_i(\epsilon) + \frac{1}{n} \sum_{i=1}^{n} \int h_i(\cdot, \epsilon) d\Delta_{iT} \tag{13}
\]
\[
0 = \frac{1}{n} \sum_{i=1}^{n} \int \frac{d^2h_i(\cdot, \epsilon)}{d\epsilon^2} dF_i(\epsilon) + 2 \frac{1}{n} \sum_{i=1}^{n} \int \frac{dh_i(\cdot, \epsilon)}{d\epsilon} d\Delta_{iT} \tag{14}
\]
\[
0 = \frac{1}{n} \sum_{i=1}^{n} \int \frac{d^3h_i(\cdot, \epsilon)}{d\epsilon^3} dF_i(\epsilon) + 3 \frac{1}{n} \sum_{i=1}^{n} \int \frac{d^2h_i(\cdot, \epsilon)}{d\epsilon^2} d\Delta_{iT} \tag{15}
\]
where \( \Delta_{iT} \equiv \sqrt{T} \left( \hat{F}_i - F_i \right) \).

**B.1 \( \theta^\epsilon (0) \)**

Because
\[
\frac{dh_i(\cdot, \epsilon)}{d\epsilon} = \frac{\partial h_i(\cdot, \epsilon)}{\partial \theta} \frac{\partial \theta}{d\epsilon} + \frac{\partial h_i(\cdot, \epsilon)}{\partial \alpha_i} \frac{\partial \alpha_i}{d\epsilon} + \frac{\partial h_i(\cdot, \epsilon)}{\partial \alpha_i} \frac{\partial \alpha_i}{d\epsilon}
\]
we may rewrite (13) as
\[
0 = \frac{1}{n} \sum_{i=1}^{n} \int \left( \frac{\partial h_i(\cdot, \epsilon)}{\partial \theta} \frac{\partial \theta}{d\epsilon} + \frac{\partial h_i(\cdot, \epsilon)}{\partial \alpha_i} \frac{\partial \alpha_i}{d\epsilon} + \frac{\partial h_i(\cdot, \epsilon)}{\partial \alpha_i} \frac{\partial \alpha_i}{d\epsilon} \right) dF_i(\epsilon) + \frac{1}{n} \sum_{i=1}^{n} \int h_i(\cdot, \epsilon) d\Delta_{iT} \tag{16}
\]
Evaluating at \( \epsilon = 0 \), and noting that
\[
E \left[ \frac{\partial U_i}{\partial \alpha_i} \right] = 0,
\]
we obtain
\[
\theta^\epsilon (0) = \left( \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \int U_i d\Delta_{iT} \right) \tag{17}
\]
We therefore have
\[
\sqrt{nT} \frac{1}{\sqrt{T}} \theta^\epsilon (0) = \left( \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i \right)^{-1} \left( \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U_i \right)
\]
\[
\xrightarrow[n \rightarrow \infty]{} N \left( 0, \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i \right)^{-1} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \Phi_i \right) \right) \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i^\prime \right)^{-1} \right)
\]

**B.2 \( \alpha_i^\theta \) and \( \alpha_i^\epsilon \)**

In the \( i \)th stratum, \( \alpha_i(\theta, F_i(\epsilon)) \) solves the estimating equation
\[
\int V_i(\cdot; \theta, \alpha_i(\theta, F_i(\epsilon))) dF_i(\epsilon) = 0 \tag{18}
\]
Differentiating the LHS with respect to \( \theta \) and \( \epsilon \), we obtain
\[
0 = \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \theta'} dF_i(\epsilon) + \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \alpha_i^\prime} dF_i(\epsilon) \right) \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \theta'},
\]
\[
0 = \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \alpha_i^\prime} dF_i(\epsilon) \right) \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \epsilon} + \int V_i(\cdot, \theta, \epsilon) d\Delta_{iT}.
\]

Observe that
\[
\frac{\partial \alpha_i(\theta, F_i(e))}{\partial \theta'} = -\left( \int \frac{\partial V_i(\cdot, \theta, e)}{\partial \alpha'_i} dF_i(e) \right)^{-1} \left( \int V_i(\cdot, \theta, e) d\Delta_T \right),
\]
\[
\frac{\partial \alpha_i(\theta, F_i(e))}{\partial e} = -\left( \int \frac{\partial V_i(\cdot, \theta, e)}{\partial \alpha'_i} dF_i(e) \right)^{-1} \left( \int V_i(\cdot, \theta, e) d\Delta_T \right).
\]

Equating these equations to zero and solving for derivatives of \( \alpha_i \) evaluated at \( e = 0 \) gives
\[
\alpha_i^\theta(0) = -\left( E \left[ \frac{\partial V_i}{\partial \alpha'_i} \right] \right)^{-1} E \left[ \frac{\partial V_i}{\partial \theta} \right] = O(1),
\]
\[
\alpha_i^\epsilon(0) = -\left( E \left[ \frac{\partial V_i}{\partial \alpha'_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T V_{it} \right) = O_p(1),
\]
where
\[
\alpha_i^\theta \equiv \frac{\partial \alpha_i(\theta, F_i(0))}{\partial \theta'}, \quad \alpha_i^\epsilon \equiv \frac{\partial \alpha_i(\theta, F_i(0))}{\partial \epsilon}.
\]

### B.3 \( \theta_{\epsilon}^\epsilon(0) \)

Because
\[
\frac{dh_i(\cdot, \epsilon)}{de} = \sum_{k=1}^p \frac{\partial h_i(\cdot, \epsilon)}{\partial \theta_k} \frac{\partial \theta_k}{\partial \epsilon} + \sum_{l=1}^q \sum_{k=1}^p \frac{\partial h_i(\cdot, \epsilon)}{\partial \alpha_l} \frac{\partial \alpha_l}{\partial \theta_k} \frac{\partial \theta_k}{\partial \epsilon} + \sum_{l=1}^q \frac{\partial h_i(\cdot, \epsilon)}{\partial \alpha_l} \frac{\partial \alpha_l}{\partial \epsilon}
\]
we have
\[
\frac{d^2 h_i(\cdot, \epsilon)}{de^2} = \sum_{k=1}^p \sum_{k'=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \theta_k \partial \theta_{k'}} \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_{k'}}{\partial \epsilon} + \sum_{l=1}^q \sum_{k=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \alpha_l \partial \theta_k} \frac{\partial \alpha_l}{\partial \theta_k} \frac{\partial \theta_k}{\partial \epsilon} + \sum_{l=1}^q \sum_{k=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \alpha_l \partial \epsilon} \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon}
\]
\[
+ \sum_{l=1}^q \sum_{l'=1}^q \sum_{k=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \alpha_l \partial \alpha_{l'}} \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \alpha_{l'}}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} + \sum_{l=1}^q \sum_{l'=1}^q \sum_{k=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \alpha_l \partial \alpha_{l'}} \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \alpha_{l'}}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon}
\]
\[
+ \sum_{l=1}^q \sum_{l'=1}^q \sum_{k=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \alpha_l \partial \alpha_{l'}} \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \alpha_{l'}}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} + \sum_{l=1}^q \sum_{l'=1}^q \sum_{k=1}^p \frac{\partial^2 h_i(\cdot, \epsilon)}{\partial \alpha_l \partial \alpha_{l'}} \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \alpha_{l'}}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon}
\]
Evaluating (14) at $\epsilon = 0$, and noting that $E[U_i^{\alpha_i}] = 0$, we obtain

$$0 = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \sum_{l=1}^{p} E \left[ \frac{\partial^2 U_i}{\partial \theta_k \partial \theta_{k'}} \right] \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_{k'}}{\partial \epsilon} + 2 \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \sum_{l=1}^{p} E \left[ \frac{\partial^2 U_i}{\partial \theta_k \partial \alpha_l} \right] \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \alpha_l}{\partial \epsilon} + \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \sum_{l=1}^{p} E \left[ \frac{\partial^2 U_i}{\partial \theta_k \partial \alpha'} \right] \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \alpha'}{\partial \epsilon}

+ \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \sum_{l=1}^{p} \left[ \left( \int \frac{\partial U_i}{\partial \theta_k} d\Delta x \right) \theta_k + \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \sum_{l=1}^{p} \left( \int \frac{\partial U_i}{\partial \alpha_l} d\Delta x \right) \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \theta_k}{\partial \epsilon} \right]

+ \frac{2}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \left( \int \frac{\partial U_i}{\partial \alpha_l} d\Delta x \right) \frac{\partial \alpha_l}{\partial \epsilon}

Because

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \sum_{l=1}^{p} E \left[ \frac{\partial^2 U_i}{\partial \alpha_l \partial \alpha'} \right] \frac{\partial \alpha_l}{\partial \epsilon} \frac{\partial \alpha'}{\partial \epsilon} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\partial \alpha}{\partial \epsilon} \frac{E \left[ \frac{\partial^2 U_i^{(1)}}{\partial \alpha_l \partial \alpha'} \right]}{\partial \epsilon} \right] = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\partial \alpha}{\partial \epsilon} \frac{E \left[ \frac{\partial^2 U_i^{(p)}}{\partial \alpha_l \partial \alpha'} \right]}{\partial \epsilon} \right] = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \frac{\partial \alpha_l}{\partial \epsilon} \frac{E \left[ \frac{\partial V_l}{\partial \alpha'} \right]}{\partial \epsilon}

= \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \frac{\partial \alpha_l}{\partial \epsilon} \frac{E \left[ \frac{\partial V_l}{\partial \alpha'} \right]}{\partial \epsilon} \right] = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \alpha_l} \frac{\partial \alpha_l}{\partial \epsilon} \right)

= \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \left( \int \frac{\partial U_i}{\partial \alpha_l} d\Delta x \right) \frac{\partial \alpha_l}{\partial \epsilon} = -2 \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \alpha_l} \right) \left( \frac{\partial V_l}{\partial \alpha'} \right) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)

2 \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{q} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \theta_k} \frac{\partial \alpha_l}{\partial \epsilon} \right) \frac{\partial \theta_k}{\partial \epsilon} = O_p \left( \frac{1}{\sqrt{n}} \right) \frac{1}{O_p \left( \frac{1}{\sqrt{n}} \right)} = O_p \left( \frac{1}{n} \right)

2 \frac{1}{n} \sum_{k=1}^{p} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \alpha_l} \frac{\partial \alpha_l}{\partial \epsilon} \right) \frac{\partial \alpha_l}{\partial \epsilon} = O_p \left( \frac{1}{\sqrt{n}} \right) \frac{1}{O_p \left( \frac{1}{\sqrt{n}} \right)} = O_p \left( \frac{1}{n} \right)

2 \frac{1}{n} \sum_{k=1}^{p} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \theta_k} \frac{\partial \theta_k}{\partial \epsilon} \right) = O_p \left( \frac{1}{\sqrt{n}} \right) \frac{1}{O_p \left( \frac{1}{\sqrt{n}} \right)} = O_p \left( \frac{1}{n} \right)

2 \frac{1}{n} \sum_{k=1}^{p} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \theta_k} \frac{\partial \alpha_l}{\partial \epsilon} \right) \frac{\partial \alpha_l}{\partial \epsilon} = O_p \left( \frac{1}{\sqrt{n}} \right) \frac{1}{O_p \left( \frac{1}{\sqrt{n}} \right)} = O_p \left( \frac{1}{n} \right)
and

$$\sum_{k=1}^{p} \sum_{k'=1}^{p} \left( \frac{1}{n} \sum_{i=1}^{n} E \left[ \frac{\partial^2 U_i}{\partial \theta_k \partial \theta_{k'}} \right] \frac{\partial \theta_k}{\partial \epsilon} \frac{\partial \theta_{k'}}{\partial \epsilon} \right) = O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( \frac{1}{n} \right) = O_p \left( \frac{1}{n} \right)$$

$$2 \sum_{l=1}^{q} \sum_{k=1}^{p} \sum_{k'=1}^{p} \left( \frac{1}{n} \sum_{i=1}^{n} E \left[ \frac{\partial^2 U_i}{\partial \theta_k \partial \alpha_{l}} \right] \frac{\partial \alpha_{l}}{\partial \epsilon} \frac{\partial \theta_{k'}}{\partial \epsilon} \right) = O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( \frac{1}{n} \right) = O_p \left( \frac{1}{n} \right)$$

$$\sum_{k=1}^{p} \sum_{k'=1}^{p} \sum_{q=1}^{q} \left( \frac{1}{n} \sum_{i=1}^{n} E \left[ \frac{\partial^2 U_i}{\partial \alpha_{l} \partial \alpha_{l'}} \right] \frac{\partial \alpha_{l}}{\partial \epsilon} \frac{\partial \alpha_{l'}}{\partial \epsilon} \right) = O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( \frac{1}{n} \right) = O_p \left( \frac{1}{n} \right)$$

we may write

$$\left( \frac{1}{n} \sum_{i=1}^{n} I_i \right) \theta^e (0) = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} E \left[ \frac{\partial^2 U_i(1)}{\partial \alpha_i \partial \alpha_i} \right] \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \right]$$

$$\vdots$$

$$\left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} E \left[ \frac{\partial^2 U_i(p)}{\partial \alpha_i \partial \alpha_i} \right] \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)$$

$$-2 \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \alpha_i} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) + O_p \left( \frac{1}{n} \right)$$

or

$$\sqrt{n} T^{1/2} \left( \frac{1}{\sqrt{T}} \right)^2 \theta^e (0)$$

$$= \frac{1}{2} \sqrt{\frac{n}{T}} \left( \frac{1}{n} \sum_{i=1}^{n} I_i \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} E \left[ \frac{\partial^2 U_i(1)}{\partial \alpha_i \partial \alpha_i} \right] \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \right]$$

$$\vdots$$

$$\left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} E \left[ \frac{\partial^2 U_i(p)}{\partial \alpha_i \partial \alpha_i} \right] \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)$$

$$- \sqrt{\frac{n}{T}} \left( \frac{1}{n} \sum_{i=1}^{n} I_i \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \alpha_i} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) + o_p (1)$$

Note that

$$\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} E \left[ \frac{\partial^2 U_i(1)}{\partial \alpha_i \partial \alpha_i} \right] \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \text{trace} \left\{ \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} E \left[ \frac{\partial^2 U_i(1)}{\partial \alpha_i \partial \alpha_i} \right] \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left[ V_i V_i' \right] \right\} + o_p (1)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial U_i}{\partial \alpha_i} \right) \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right) = \frac{1}{n} \sum_{i=1}^{n} \text{trace} \left\{ \left( E \left[ \frac{\partial V_i}{\partial \alpha_i} \right] \right)^{-1} \left[ V_i \frac{\partial U_i(1)}{\partial \alpha_i} \right] \right\} + o_p (1)$$

It therefore follows that

$$\sqrt{n} T^{1/2} \left( \frac{1}{\sqrt{T}} \right)^2 \theta^e (0) = \frac{1}{2} \sqrt{\frac{n}{T}} \left( \frac{1}{n} \sum_{i=1}^{n} I_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i \right) + o_p (1)$$

13
B.4 $\alpha_i^{\theta\theta}$, $\alpha_i^{\theta\varepsilon}$, and $\alpha_i^{\varepsilon\varepsilon}$

Second order differentiation \( \left( \frac{\partial^2}{\partial \theta \partial \theta'}, \frac{\partial^2}{\partial \varepsilon \partial \varepsilon'}, \frac{\partial^2}{\partial \varepsilon \partial \theta'} \right) \) of (18) yields

\[
0 = \int \frac{\partial^2 V_i(\cdot, \theta, \epsilon)}{\partial \theta \partial \theta'} dF_i(\epsilon) + \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \theta} \left( \int \frac{\partial^2 V_i(\cdot, \theta, \epsilon)}{\partial \theta \partial \theta'} dF_i(\epsilon) \right) + \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \theta} dF_i(\epsilon) \right) \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial \theta \partial \theta'} + \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \theta} dF_i(\epsilon) \right) \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \theta} \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \theta},
\]

and

\[
0 = \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \alpha_i} dF_i(\epsilon) \right) \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial \epsilon^2} + \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \alpha_i} dF_i(\epsilon) \right) \left( \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \epsilon} \right)^2 + 2 \left( \int \frac{\partial V_i(\cdot, \theta, \epsilon)}{\partial \alpha_i} dF_i(\epsilon) \right) \frac{\partial \alpha_i(\theta, F_i(\epsilon))}{\partial \epsilon},
\]

These three equalities characterize \( \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial \theta \partial \theta'} \), \( \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial \varepsilon \partial \varepsilon'} \), and \( \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial \varepsilon \partial \theta'} \).

**Lemma 4**

\[
T \Pr \left[ \max_{0 \leq \epsilon \leq \sqrt{T}} \left| \hat{\theta}(\epsilon) - \theta_0 \right| \geq \eta \right] = o(1)
\]

and

\[
T \Pr \left[ \max_{1 \leq i \leq n} \max_{0 \leq \epsilon \leq \sqrt{T}} \left| \hat{\alpha}_i(\epsilon) - \alpha_{i0} \right| \geq \eta \right] = o(1)
\]

for every \( \eta > 0 \).

**Proof.** Only the first assertion is proved. The second assertion can be proved similarly. Let \( \eta \) be given, and let \( \epsilon \equiv \inf_i \inf_{\{(\theta, \alpha);\alpha-(\theta, \alpha)\geq \eta\}} |G_{(i)}(\theta, \alpha)| > 0 \). Recall that

\[
F(\epsilon) \equiv F + \epsilon \sqrt{T} \left( \bar{F} - F \right), \quad \epsilon \in \left[ 0, \frac{1}{\sqrt{T}} \right]
\]

We have

\[
\int g(\cdot, \theta, \alpha_i(\theta)) dF_i(\epsilon) = \left( 1 - \epsilon \sqrt{T} \right) G_{(i)}(\theta, \alpha_i) + \epsilon \sqrt{T} \bar{G}_{(i)}(\theta, \alpha_i)
\]

and

\[
\int g(\cdot, \theta, \alpha_i(\theta)) dF_i(\epsilon) - G_{(i)}(\theta, \alpha_i) \leq \left( 1 - \epsilon \sqrt{T} \right) \left| \bar{G}_{(i)}(\theta, \alpha) - G_{(i)}(\theta, \alpha) \right| \leq \left| \bar{G}_{(i)}(\theta, \alpha) - G_{(i)}(\theta, \alpha) \right|
\]

14
By Lemma 3, we have

\[
\Pr \left[ \max_{0 \leq t \leq \frac{1}{\sqrt{T}}} \max_{1 \leq i \leq n} \sup_{(\theta, \alpha)} \left| \int g(\cdot; \theta, \alpha_i(\theta)) \, dF_i(\epsilon) - G_i(\theta, \alpha) \right| \geq \eta \right] = o(T^{-1})
\]

Therefore, for every \(0 \leq \epsilon \leq \frac{1}{\sqrt{T}}\) with probability equal to \(1 - o(T^{-1})\), we have

\[
\min_{|\theta - \theta_0| > \eta, \alpha_1, \ldots, \alpha_n} n^{-1} \sum_{i=1}^{n} \left| \int g(\cdot; \theta, \alpha_i(\theta)) \, dF_i(\epsilon) \right| \geq \min_{|\theta, \alpha| > (\theta_0, \alpha_0)} n^{-1} \sum_{i=1}^{n} \left| \int g(\cdot; \theta, \alpha_i(\theta)) \, dF_i(\epsilon) \right|
\]

\[
> \min_{|\theta, \alpha| > (\theta_0, \alpha_0)} n^{-1} \sum_{i=1}^{n} \left| G_i(\theta, \alpha_i) \right| - \frac{1}{2} \epsilon
\]

\[
> \frac{1}{2} \epsilon
\]

\[
= n^{-1} \sum_{i=1}^{n} \left| \int g(\cdot; \hat{\theta}(\epsilon), \alpha_i(\hat{\theta}(\epsilon))) \, dF_i(\epsilon) \right|
\]

We therefore obtain that \(\Pr \left[ \max_{0 \leq t \leq \frac{1}{\sqrt{T}}} \left| \hat{\theta}(\epsilon) - \theta_0 \right| \geq \eta \right] = o(T^{-1})\).

**Lemma 5** Suppose that, for each \(i\), \(\{\xi_{it}(\phi), t = 1, 2, \ldots\}\) is a sequence of zero mean i.i.d. random variables indexed by some parameter \(\phi \in \Phi\). We assume that \(\{\xi_{it}(\phi), t = 1, 2, 3\}\) are independent across \(i\). We also assume that \(\sup_{\phi \in \Phi} |\xi_{it}(\phi)| \leq B_{it}\) for some sequence of random variables \(B_{it}\) that is i.i.d. across \(t\) and independent across \(i\). Finally, we assume that \(\max_i E \left[ |B_{it}|^{64} \right] < \infty\), and \(n = O(T)\). We then have

\[
\Pr \left[ \max_{i} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_{it}(\phi_i) > T^{1-v} \right] = o \left( \frac{1}{T} \right)
\]

for every \(v\) such that \(0 \leq v < \frac{11}{160}\). Here, \(\{\phi_i\}\) is an arbitrary sequence in \(\Phi\).

**Proof.** By Markov’s inequality,

\[
\Pr \left[ \sup_{\phi \in \Phi} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_{it}(\phi_i) \right| > T^{1-v} \right] = \Pr \left[ \sup_{\phi \in \Phi} \left| \sum_{t=1}^{T} \xi_{it}(\phi_i) \right| > T^{1-v} \right]
\]

\[
\leq E \left[ \sup_{\phi \in \Phi} \left| \sum_{t=1}^{T} \xi_{it}(\phi_i) \right|^{64} \right] T^{32} = \sup_{\phi \in \Phi} E \left[ \sum_{t=1}^{T} \xi_{it}(\phi_i) \right]^{64} T^{32} \left| \sum_{t=1}^{T} \xi_{it}(\phi_i) \right|^{64}
\]

where the last equality is based on dominated convergence. By Lemma 1, we have

\[
E \left[ \sum_{t=1}^{T} \xi_{it}(\phi_i) \right]^{64} \leq C T^{32} \cdot E \left[ |\xi_{it}(\phi_i)|^{64} \right] \leq T^{32} \cdot C \max_i E \left[ |B_{it}|^{64} \right]
\]

for some \(C\). Therefore, we have

\[
T^2 \Pr \left[ \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_{it}(\phi_i) \right| > T^{1-v} \right] \leq T^2 \frac{CT^{32}}{T^{128} - 64v \frac{1}{64}} = O \left( T^{-11 + 64v} \right),
\]

15
Therefore, we have

$$T \Pr \left[ \max_i \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_{it} (\phi_i) \right| > T^{1 \over \alpha} - v \right] \leq T \sum_{i=1}^{n} \Pr \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_{it} (\phi_i) \right] > T^{1 \over \alpha} - v \right] = nT \cdot \frac{C T^{32}}{T^{1 \over \alpha} - 64v^{\alpha} T} = o(1).$$

\*\*\* 

**Lemma 6** Suppose that $K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon))$ is equal to

$$\frac{\partial^{m_1 + m_2} g (X_i; \theta (e), \alpha_i (\theta (e), \epsilon))}{\partial \theta^{m_1} \partial \alpha_i^{m_2}}$$

for some $m_1 + m_2 \leq 1, \ldots, 5$. Then, for any $\eta > 0$, we have

$$\Pr \left[ \max_{0 \leq \epsilon \leq 1 \over \alpha} \left| \frac{1}{n} \sum_{i=1}^{n} \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) dF_i (e) - \frac{1}{n} \sum_{i=1}^{n} E [K_i (X_i; \theta_0, \alpha_i)] \right| > \eta \right] = o \left( T^{-1} \right)$$

and

$$\Pr \left[ \max_{0 \leq \epsilon \leq 1 \over \alpha} \left| \frac{1}{n} \sum_{i=1}^{n} \left( \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) dF_i (e) - E [K_i (X_i; \theta_0, \alpha_i)] \right) > \eta \right] = o \left( T^{-1} \right).$$

Also,

$$\Pr \left[ \max_{0 \leq \epsilon \leq 1 \over \alpha} \left| \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) d\Delta_i T \right| > C T^{1 \over \alpha} - v \right] = o \left( T^{-1} \right)$$

for some constant $C > 0$ and $0 \leq v < 11 \over 100$.

**Proof.** Note that we may write

$$\left\| \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) dF_i (e) - \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) dF_i \right\|$$

$$\leq \left\| \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) dF_i (e) - \int K_i (\cdot; \theta (0), \alpha_i (\theta (0), 0)) dF_i (e) \right\|$$

$$\quad + \left\| \int K_i (\cdot; \theta (0), \alpha_i (\theta (0), 0)) dF_i (e) - \int K_i (\cdot; \theta (0), \alpha_i (\theta (0), 0)) dF_i \right\|$$

$$\leq \int M(X_i) (\| \theta (e) - \theta \| + | \alpha_i (\theta (e), \epsilon) - \alpha_i |) dF_i (e)$$

$$\quad + \epsilon \sqrt{T} \left\| \int K_i (\cdot; \theta (0), \alpha_i (\theta (0), 0)) d \left( F_i - F_i \right) \right\|.$$

Therefore, we have

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \int K_i (\cdot; \theta (e), \alpha_i (\theta (e), \epsilon)) dF_i (e) - \frac{1}{n} \sum_{i=1}^{n} E [K_i (X_i; \theta_0, \alpha_i)] \right\|$$

$$\leq \| \theta (e) - \theta \| \cdot \left\| \frac{1}{n} \sum_{i=1}^{n} \left( E [M(X_i)] + \frac{T}{T} \sum_{t=1}^{T} M(X_i) \right) \right\|$$

$$\quad + \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( \alpha_i (\theta (e), \epsilon) - \alpha_i \right)^2 \right\}^{1 \over 2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( E [M(X_i)] + \frac{T}{T} \sum_{t=1}^{T} M(X_i) \right) \right\}^{1 \over 2}$$

$$\quad + \left\| \frac{1}{n} \sum_{i=1}^{n} \left( \frac{T}{T} \sum_{t=1}^{T} K_i (X_i; \theta (0), \alpha_i (\theta (0), 0)) - E [K_i (X_i; \theta (0), \alpha_i (\theta (0), 0))] \right) \right\|.$$
the RHS of which can be bounded by using Lemmas ?? and 4 in absolute value by some \( \eta > 0 \) with probability \( 1 - o(T^{-1}) \).

Because

\[
\left| \int K_i (\cdot; \theta (\epsilon), \alpha_i (\theta (\epsilon), \epsilon)) dF_i (\epsilon) - E [K_i (X; \theta_0, \alpha_0)] \right|
\leq |\theta (\epsilon) - \theta| \cdot \left( E [M (X_{it})] + \frac{1}{T} \sum_{t=1}^{T} M (X_{it}) \right)
+ |\alpha_i (\theta (\epsilon), \epsilon) - \alpha_i| \cdot \left( E [M (X_{it})] + \frac{1}{T} \sum_{t=1}^{T} M (X_{it}) \right)
+ \left| \frac{1}{T} \sum_{t=1}^{T} M (X_{it}) - E [M (X_{it})] \right|
\]

we can bound

\[
\max_i \max_{0 \leq \epsilon \leq \frac{1}{\sqrt{T}}} \left| \int K_i (\cdot; \theta (\epsilon), \alpha_i (\theta (\epsilon), \epsilon)) dF_i (\epsilon) - E [K_i (X; \theta_0, \alpha_0)] \right|
\]

in absolute value by some \( \eta > 0 \) with probability \( 1 - o(T^{-1}) \).

Using Lemma 5, we can also show that

\[
\max_i \left| \int K_i (\cdot; \theta (\epsilon), \alpha_i (\theta (\epsilon), \epsilon)) d\Delta_iT \right|
\]

can be bounded by in absolute value by \( CT^{\frac{1}{2} - v} \) for some constant \( C > 0 \) and \( v \) such that \( 0 \leq v < \frac{11}{160} \) with probability \( 1 - o(T^{-1}) \). ■

**Lemma 7**

\[
\Pr \left[ \max_i \max_{0 \leq \epsilon \leq \frac{1}{\sqrt{T}}} \left| \alpha_i^\theta (\epsilon) \right| > C \right] = o(T^{-1})
\]

\[
\Pr \left[ \max_i \max_{0 \leq \epsilon \leq \frac{1}{\sqrt{T}}} \left| \alpha_i^\epsilon (\epsilon) \right| > CT^{\frac{1}{2} - v} \right] = o(T^{-1})
\]

for some constant \( C > 0 \) and \( 0 \leq v < \frac{11}{160} \).

**Proof.** From Appendix B.2, we obtain

\[
\frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} = - \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} dF_i (\epsilon) \right)^{-1} \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \theta} dF_i (\epsilon) \right),
\]

\[
\frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon} = - \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} dF_i (\epsilon) \right)^{-1} \left( \int V_i (\cdot, \theta, \epsilon) d\Delta_iT \right).
\]

Using Lemma 6, we can see that \( \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} dF_i (\epsilon) \right)^{-1} \) is uniformly bounded away from zero with probability \( 1 - o(T^{-1}) \). We can also see that, with probability \( 1 - o(T^{-1}) \), \( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \epsilon} dF_i (\epsilon) \) is uniformly bounded by some constant \( C \), and \( \int V_i (\cdot, \theta, \epsilon) d\Delta_iT \) is uniformly bounded by \( CT^{\frac{1}{2} - v} \). ■
Lemma 8

\[ \Pr \left[ \max_{0 \leq t \leq \frac{\tau}{T}} |\theta^\epsilon (t)| > CT^{\frac{1}{4v} - v} \right] = o \left( T^{-1} \right) \]

for some constant \( C > 0 \) and \( 0 \leq v < \frac{11}{100} \).

**Proof.** From (16), we have

\[ \theta^\epsilon (t) = - \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \int \left( \frac{\partial h_i (\cdot, t)}{\partial \theta} + \frac{\partial h_i (\cdot, t)}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial \theta} \right) \, dF_i (\epsilon) \right) \right] - \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \alpha_i}{\partial \epsilon} \left( \int \frac{\partial h_i (\cdot, t)}{\partial \alpha_i} \, dF_i (\epsilon) \right) + \frac{1}{n} \sum_{i=1}^{n} \int h_i (\cdot, t) \, d\Delta x_T \]

Using Lemmas 6, and 7, we can bound the denominator of \( \theta^\epsilon (t) \) by some \( C > 0 \), and the numerator by some \( CT^{\frac{1}{4v} - v} \) with probability \( 1 - o \left( T^{-1} \right) \).

Lemma 9

\[ \Pr \left[ \max_{0 \leq t \leq \frac{\tau}{T}} |\alpha^\theta_{t, t, i} (\epsilon)| > C \right] = o \left( T^{-1} \right) \]

\[ \Pr \left[ \max_{0 \leq t \leq \frac{\tau}{T}} |\alpha^\epsilon_{t, t, i} (\epsilon)| > C T^{\frac{1}{4v} - v} \right] = o \left( T^{-1} \right) \]

\[ \Pr \left[ \max_{0 \leq t \leq \frac{\tau}{T}} |\alpha^\epsilon_{t, t, i} (\epsilon)| > C \left( T^{\frac{1}{4v} - v} \right)^2 \right] = o \left( T^{-1} \right) \]

for some constant \( C > 0 \) and \( 0 \leq v < \frac{11}{100} \). Here, \( \alpha^\theta_{t, t, i} (\epsilon) \equiv \frac{\partial^2 \alpha_i}{\partial \theta \partial \theta^T} \). We similarly define \( \alpha^\epsilon_{t, t, i} \).

**Proof.** Note that

\[ 0 = \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \theta \partial \theta^T} \, dF_i (\epsilon) + \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \theta^T} \, dF_i (\epsilon) \right) + \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} \, dF_i (\epsilon) \right) \frac{\partial^2 \alpha_i (\theta, F_i (\epsilon))}{\partial \theta \partial \theta^T} + \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} \, dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta^T}, \]

\[ 0 = \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \theta \partial \alpha_i} \, dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} + \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} \, dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta^T}, \]

and

\[ 0 = \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} \, dF_i (\epsilon) \right) \frac{\partial^2 \alpha_i (\theta, F_i (\epsilon))}{\partial \theta \partial \theta^T} + \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} \, dF_i (\epsilon) \right) \partial \alpha_i (\theta, F_i (\epsilon)) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta^T} + 2 \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i} \, d\Delta x_T \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta}. \]

The result then follows by applying the same argument as in the proof of Lemma 7. ■
**Lemma 10**

\[
\Pr \left[ \max_{0 \leq \varepsilon \leq \frac{1}{T}} |\theta^{\varepsilon}(\varepsilon)| > C \left( T^{\frac{1}{2}} - \nu \right)^2 \right] = o(T^{-1})
\]

for some constant \( C > 0 \) and \( 0 \leq \nu < \frac{11}{160} \).

**Proof.** The conclusion follows by using the characterization of \( \theta^{\varepsilon}(\varepsilon) \) in Appendix B.3, and Lemmas 6, 7, 8, and 9. \( \blacksquare \)

**Lemma 11**

\[
\Pr \left[ \max_{i} \max_{0 \leq \varepsilon \leq \frac{1}{T}} |\alpha_i^{\theta_{i},\theta_{i}\nu}(\varepsilon)| > C \right] = o(T^{-1})
\]

\[
\Pr \left[ \max_{i} \max_{0 \leq \varepsilon \leq \frac{1}{T}} |\alpha_i^{\theta_{i},\theta_{i}\nu}(\varepsilon)| > C T^{\frac{1}{2}}\nu \right] = o(T^{-1})
\]

\[
\Pr \left[ \max_{i} \max_{0 \leq \varepsilon \leq \frac{1}{T}} |\alpha_i^{\theta_{i},\theta_{i}\nu}(\varepsilon)| > C \left( T^{\frac{1}{2}} - \nu \right)^2 \right] = o(T^{-1})
\]

\[
\Pr \left[ \max_{i} \max_{0 \leq \varepsilon \leq \frac{1}{T}} |\alpha_i^{\theta_{i},\theta_{i}\nu}(\varepsilon)| > C \left( T^{\frac{1}{2}} - \nu \right)^3 \right] = o(T^{-1})
\]

for some constant \( C > 0 \) and \( 0 \leq \nu < \frac{11}{160} \).

**Proof.** It was seen in Appendix B.4 that

\[
0 = \left( \int \frac{\partial^2 V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \theta, \partial \alpha_i} \right) \frac{\partial \alpha_i(\theta, F_i(\varepsilon))}{\partial \varepsilon} + \left( \int \frac{\partial V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \alpha_i} \right) \frac{\partial^2 \alpha_i(\theta, F_i(\varepsilon))}{\partial \theta, \partial \varepsilon} + \left( \int \frac{\partial V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \alpha_i} \right) \frac{\partial^2 \alpha_i(\theta, F_i(\varepsilon))}{\partial \theta, \partial \varepsilon} + \left( \int \frac{\partial V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \alpha_i} \right) \frac{\partial \alpha_i(\theta, F_i(\varepsilon))}{\partial \theta, \partial \varepsilon},
\]

and

\[
0 = \left( \int \frac{\partial V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \alpha_i} \right) \frac{\partial^2 \alpha_i(\theta, F_i(\varepsilon))}{\partial \varepsilon^2} + \left( \int \frac{\partial^2 V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \alpha_i^2} \right) \frac{\partial \alpha_i(\theta, F_i(\varepsilon))}{\partial \varepsilon} + \frac{2 \left( \int \frac{\partial V_i(\cdot, \theta, \varepsilon) \, dF_i(\varepsilon)}{\partial \alpha_i} \right) \frac{\partial \alpha_i(\theta, F_i(\varepsilon))}{\partial \varepsilon}}{2}.
\]
We therefore obtain
\[
0 = \int \frac{\partial^3 V_i (\cdot, \theta, \epsilon)}{\partial \theta_r \partial \theta_r \partial \theta_r} dF_i (\epsilon) + \left( \int \frac{\partial^3 V_i (\cdot, \theta, \epsilon)}{\partial \theta_r \partial \theta_r \partial \theta_r} dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \\
+ \frac{\partial^2 \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i \partial \theta_r} dF_i (\epsilon) \right) + \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^3 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i \partial \theta_r \partial \theta_r} dF_i (\epsilon) \right) \\
+ \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i^2} dF_i (\epsilon) \right) + \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i^2} dF_i (\epsilon) \right) \\
+ \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i^2} dF_i (\epsilon) \right) + \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i^2} dF_i (\epsilon) \right) \\
+ \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i^2} dF_i (\epsilon) \right) + \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \theta_r} \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \alpha_i^2} dF_i (\epsilon) \right) ,
\]
which characterizes $\frac{\partial^3 \alpha_i(\theta,F_i(\epsilon))}{\partial \theta_i \partial \theta_{i'} \partial \epsilon}$,

\[
0 = \left( \int \frac{\partial^3 V_i (\cdot, \theta, \epsilon)}{\partial \theta_i \partial \theta_{i'} \partial \epsilon} dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon} + \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \theta_i \partial \epsilon} dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon} + \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \epsilon} dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon}
\]

which characterizes $\frac{\partial^3 \alpha_i(\theta,F_i(\epsilon))}{\partial \theta_i \partial \theta_{i'} \partial \epsilon}$,

\[
0 = \left( \int \frac{\partial^2 V_i (\cdot, \theta, \epsilon)}{\partial \theta_i \partial \epsilon} dF_i (\epsilon) \right) \frac{\partial^2 \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon^2} + \left( \int \frac{\partial V_i (\cdot, \theta, \epsilon)}{\partial \epsilon} dF_i (\epsilon) \right) \frac{\partial \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon} + \left( \int \frac{\partial^2 \alpha_i (\theta, F_i (\epsilon))}{\partial \epsilon^2} dF_i (\epsilon) \right)
\]
which characterizes $\frac{\partial^3 \alpha_i(\theta, F_i(e))}{\partial \theta \partial \theta ^2}$, and

$$
0 = \left( \int \frac{\partial^2 V_i(\cdot, \theta, e)}{\partial \alpha_i^2} dF_i(e) \right) \frac{\partial \alpha_i(\theta, F_i(e))}{\partial \theta} \frac{\partial^2 \alpha_i(\theta, F_i(e))}{\partial \theta^2} + \left( \int \frac{\partial V_i(\cdot, \theta, e)}{\partial \alpha_i} \frac{\partial^2 \alpha_i(\theta, F_i(e))}{\partial \theta^2} d\Delta_{iT} \right) \frac{\partial^2 \alpha_i(\theta, F_i(e))}{\partial \theta^2} \\
+ \left( \int \frac{\partial^3 V_i(\cdot, \theta, e)}{\partial \alpha_i^3} dF_i(e) \right) \frac{\partial^3 \alpha_i(\theta, F_i(e))}{\partial \theta^3}
$$

Combining Lemmas 6, 7, 8, 9, and 10, we can bound $\frac{1}{n} \sum_{i=1}^{n} \int \frac{d^2 h_i(\cdot, e)}{d^2 \theta} d\Delta_{iT}$ by $C \left( T^{1/3 - v} \right)$ with probability $1 - o(T^{-1})$. It was seen in Appendix B.3 that the $r$-th component of $\frac{d^2 h_i(\cdot, e)}{d^2 \theta}$ is equal to

$$
\frac{d^2 h_i^{(r)}(\cdot, e)}{d^2 \theta} = \frac{\partial \theta(\cdot, e)}{\partial \theta} \frac{\partial^2 h_i^{(r)}(\cdot, e)}{\partial \theta \partial \theta} \frac{\partial \theta(\cdot, e)}{\partial \theta} + \frac{\partial^2 h_i^{(r)}(\cdot, e)}{\partial \theta \partial \theta} \frac{\partial \alpha_i(\cdot, e)}{\partial \theta} \frac{\partial \theta(\cdot, e)}{\partial \theta} + \frac{\partial^2 h_i^{(r)}(\cdot, e)}{\partial \theta \partial \theta} \frac{\partial \alpha_i(\cdot, e)}{\partial \theta} \frac{\partial \alpha_i(\cdot, e)}{\partial \theta} \\
+ \frac{\partial^2 h_i^{(r)}(\cdot, e)}{\partial \theta \partial \theta} \frac{\partial \alpha_i(\cdot, e)}{\partial \theta} \frac{\partial ^2 \theta}{\partial \theta^2} + \frac{\partial h_i^{(r)}(\cdot, e)}{\partial \theta} \frac{\partial ^2 \theta}{\partial \theta^2} + \frac{\partial h_i^{(r)}(\cdot, e)}{\partial \theta} \frac{\partial ^2 \theta}{\partial \theta^2}
$$

Using Lemmas 6, 7, 8, 9, and 10 again, we can conclude that $\frac{1}{n} \sum_{i=1}^{n} \int \frac{d^2 h_i(\cdot, e)}{d^2 \theta} d\Delta_{iT}$ plus terms that can all be bounded by $\frac{1}{n} \sum_{i=1}^{n} \int \frac{d^2 h_i(\cdot, e)}{d^2 \theta} d\Delta_{iT}$ by $C \left( T^{1/3 - v} \right)$.
with probability $1 - o(T^{-1})$. Because $\left( \frac{1}{n} \sum_{i=1}^{n} \int \frac{\partial h_i(\cdot, \epsilon)}{\partial \theta} dF_i(\epsilon) \right)^{-1}$ is bounded away from 0 by Lemma 6, we obtain the desired conclusion.

Lemma 13

$$\sqrt{nT} (\hat{\theta} - \theta_0) = \sqrt{nT} \frac{1}{\sqrt{T}} \theta^\epsilon (0) + \sqrt{nT} \frac{1}{2} \left( \frac{1}{\sqrt{T}} \right)^2 \theta^{\epsilon^2} (0) + o_p (1)$$

Proof. Follows from equation (10) and Lemma 12.

C Proof of Theorem 1

Let $\tilde{\alpha}_i (\epsilon)$ be such that

$$0 = \int V_i \left( \cdot, \hat{\theta}(\epsilon), \hat{\alpha}_i (\epsilon) \right) dF_i (\epsilon)$$

(Note slight difference of definition. In previous sections, $\tilde{\alpha}_i (\theta, \epsilon)$ was understood to be a solution to $\int V_i (\cdot; \theta, \alpha_i (\theta, F_i (\epsilon))) dF_i (\epsilon) = 0$.) Using the same arguments as earlier, we are looking for the expansion

$$\tilde{\alpha}_i (\epsilon) - \alpha_{i0} = \frac{1}{\sqrt{T}} \hat{\alpha}_i (0) + \frac{1}{2T} \hat{\alpha}_i^\epsilon (\hat{\epsilon})$$

for some $\hat{\epsilon} \in \left[ 0, T^{-1/2} \right]$. Let

$$v_i (\cdot, \epsilon) \equiv V_i (\theta (F (\epsilon)), \alpha_i (F_i (\epsilon))) .$$

The first order condition may be written as

$$0 = \int v_i (\cdot, \epsilon) dF_i (\epsilon)$$

Differentiating repeatedly with respect to $\epsilon$, we obtain

$$0 = \int \frac{dv_i (\cdot, \epsilon)}{d\epsilon} dF_i (\epsilon) + \int v_i (\cdot, \epsilon) d\Delta_{iT}$$

(21)

$$0 = \int \frac{d^2 v_i (\cdot, \epsilon)}{d\epsilon^2} dF_i (\epsilon) + 2 \int \frac{dv_i (\cdot, \epsilon)}{d\epsilon} d\Delta_{iT}$$

(22)

where $\Delta_{iT} \equiv \sqrt{T} (\hat{F}_i - F_i)$. Because

$$\frac{dv_i (\cdot, \epsilon)}{d\epsilon} = \frac{\partial v_i (\cdot, \epsilon)}{\partial \theta} \frac{\partial \theta}{d\epsilon} + \frac{\partial v_i (\cdot, \epsilon)}{\partial \alpha_i} \frac{\partial \alpha_i}{d\epsilon}$$

we may rewrite (21) as

$$0 = \int \left( \frac{\partial v_i (\cdot, \epsilon)}{\partial \theta} \frac{\partial \theta}{d\epsilon} + \frac{\partial v_i (\cdot, \epsilon)}{\partial \alpha_i} \frac{\partial \alpha_i}{d\epsilon} \right) dF_i (\epsilon) + \int v_i (\cdot, \epsilon) d\Delta_{iT}$$

Evaluating at $\epsilon = 0$ we obtain

$$\frac{\partial v_i (\cdot, \epsilon)}{\partial \theta} (0) = -E \left[ V_i^\alpha \theta^\epsilon \right] \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_i (\cdot, \theta_0, \alpha_{i0}) + E \left[ V_i^\theta \theta^\epsilon (0) \right] \right)$$

(23)
where \( \theta^\varepsilon (0) \) is defined in (17). It also follows that
\[
\tilde{\alpha}_i^\varepsilon (e) = - \left( \int \frac{\partial v_i (\cdot, e)}{\partial \alpha_i} dF_i (e) \right)^{-1} \left[ \int \left( \frac{\partial v_i (\cdot, e)}{\partial \theta^\varepsilon} \right) dF_i (e) \theta^\varepsilon (e) + \int v_i (\cdot, e) d\Delta_{iT} \right].
\]  
(24)

Next, consider
\[
\frac{d^2 v_i (\cdot, e)}{d\varepsilon^2} = \frac{\partial^2 v_i (\cdot, e)}{\partial \varepsilon \partial \theta^\varepsilon} \frac{\partial \theta^\varepsilon}{\partial \varepsilon} + \frac{\partial^2 v_i (\cdot, e)}{\partial \theta^\varepsilon^2} \left( \frac{\partial \varepsilon}{\partial \alpha_i} \right)^2 \left( \frac{\partial \varepsilon}{\partial \alpha_i} \right)^2 + \frac{\partial^2 v_i (\cdot, e)}{\partial \alpha_i \partial \varepsilon} \frac{\partial \varepsilon}{\partial \varepsilon} + \frac{\partial^2 v_i (\cdot, e) \partial^2 \alpha_i}{\partial \varepsilon^2}
\]
such that \( \tilde{\alpha}_i^\varepsilon (e) \) is characterized by
\[
0 = \theta^\varepsilon (e) \int \frac{\partial v_i (\cdot, e)}{\partial \theta^\varepsilon} dF_i (e) \theta^\varepsilon (e) + \int \frac{\partial v_i (\cdot, e)}{\partial \theta^\varepsilon} dF_i (e) \theta^\varepsilon (e) \tilde{\alpha}_i^\varepsilon (e) + \int \frac{\partial^2 v_i (\cdot, e)}{\partial \varepsilon \partial \alpha_i} dF_i (e) \left( \tilde{\alpha}_i^\varepsilon (e) \right)^2
\]
\[
+ \int \frac{\partial v_i (\cdot, e)}{\partial \alpha_i} dF_i (e) \tilde{\alpha}_i^\varepsilon (e) + \int \frac{\partial v_i (\cdot, e)}{\partial \alpha_i} dF_i (e) \Delta_{iT} \theta^\varepsilon (e) + \int \frac{\partial v_i (\cdot, e)}{\partial \alpha_i} d\Delta_{iT} \tilde{\alpha}_i (e)
\]  
(25)

We now show that
\[
\Pr \left[ \max_i \sqrt{T} (\tilde{\alpha}_i - \alpha_{i0}) > T^{1/4 - \nu} \right] = o \left( T^{-1} \right),
\]
for any \( 0 \leq \nu < \frac{1}{160} \). For this purpose, it suffices to prove
\[
\Pr \left[ \sup_{\varepsilon \in [0,1/\sqrt{T}]} \left| \tilde{\alpha}_i^\varepsilon (e) \right| > T^{1/4 - \nu} \right] = o \left( T^{-1} \right),
\]
\[
\Pr \left[ \max_i \left| \tilde{\alpha}_i^\varepsilon (0) \right| > T^{1/4 - \nu} \right] = o \left( T^{-1} \right),
\]
\[
\Pr \left[ \max_i \sup_{\varepsilon \in [0,1/\sqrt{T}]} \left| \tilde{\alpha}_i^\varepsilon (e) \right| > \left( T^{1/4 - \nu} \right)^2 \right] = o \left( T^{-1} \right).
\]

In order to prove the first claim, we note that
\[
\Pr \left[ \sup_{\varepsilon \in [0,1/\sqrt{T}]} \left| \theta^\varepsilon (e) \right| \geq T^{1/4 - \nu} \right] = o \left( T^{-1} \right)
\]
from Lemma (8). By Lemma (6), we also have
\[
\Pr \left[ \max_i \sup_{\varepsilon \in [0,1/\sqrt{T}]} \left| \int \left( \frac{\partial v_i (\cdot, e)}{\partial \theta^\varepsilon} \right) dF_i (e) - E \left[ \frac{\partial v_i (\cdot, e)}{\partial \theta^\varepsilon} \right] \right| > \eta \right] = o \left( T^{-1} \right),
\]
\[
\Pr \left[ \max_i \sup_{\varepsilon \in [0,1/\sqrt{T}]} \left| \int \frac{\partial v_i (\cdot, e)}{\partial \alpha_i} dF_i (e) - E \left[ \frac{\partial v_i (\cdot, e)}{\partial \alpha_i} \right] \right| > \eta \right] = o \left( T^{-1} \right).
\]
By Lemma (6) again, it follows that
\[
\Pr \left[ \max_i \sup_{\varepsilon \in [0,1/\sqrt{T}]} \left| \int v_i (\cdot, e) d\Delta_{iT} \right| > T^{1/10 - \nu} \right] = o \left( T^{-1} \right).
\]
This proves the result for $\hat{\alpha}_i^\varepsilon (\epsilon)$ as well as $\hat{\alpha}_i(0)$. For $\hat{\alpha}_i^{\varepsilon r} (\epsilon)$, the result follows from representation (25) as well as Lemmas (6), (8), and (10).

We now prove Theorem 1. For this purpose, we combine

\[
\hat{\alpha}_i (\epsilon) = \alpha_{i0} + \frac{1}{\sqrt{T}} \hat{\alpha}_i^\varepsilon (0) + \frac{1}{2T} \hat{\alpha}_i^{\varepsilon r} (\epsilon),
\]

\[
\hat{\alpha}_i^r (0) = - (E [V_{i0}])^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T v_i (\cdot, \theta_0, \alpha_{i0}) + E [V_i^\theta] \theta^r (0) \right)
\]

\[
\theta^r (0) = \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^n I_i \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=1}^T U_i \right)
\]

and obtain

\[
\sqrt{T} (\hat{\alpha}_i - \alpha_{i0}) = - (E [V_{i0}])^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T v_i (\cdot, \theta_0, \alpha_{i0}) \right)
\]

\[
- \frac{1}{\sqrt{n}} (E [V_{i0}])^{-1} E [V_i^\theta] \left( \frac{1}{n} \sum_{i=1}^n I_i \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=1}^T U_i \right)
\]

\[
+ \frac{1}{2\sqrt{T}} \hat{\alpha}_i^{\varepsilon r} (\epsilon),
\]

from which the conclusion follows.
References


