Supplement to "Factor-Driven Two-Regime Regression"

Sokbae Lee
Columbia University

Yuan Liao
Rutgers University

Myung Hwan Seo
Seoul National University

Youngki Shin
McMaster University

September 7, 2020

Contents

A Identification

In this section, we establish sufficient conditions under which $(\beta_0', \delta_0', \gamma_0')'$ is identified. Recall that the covariates x_t and f_t may not be directly observable in our general setup; however, since we assume that they can be consistently estimable, it suffices to consider the identification of the unknown parameters under the simple setup that x_t and f_t are observed directly from the data.

If there is no random variable in f_t with a non-zero coefficient, γ_0 is unidentifiable. Assumption 2.1 in the main text avoids this directly by assuming that the first coefficient of γ_0 is 1.[1] We partition $f_t = (f_{1t}, f_{2t}')'$ and $\gamma = (1, \gamma_2')'$, and write, occasionally, $1\{f_{1t} > f_{2t}'\gamma\}$ instead of $1\{f_t'\gamma > 0\}$.

Remark A.1 (Alternative Scale Normalization)**.** We may consider an alternative parameter space for γ_0: $\gamma_0 \in \Gamma \equiv \{\gamma : |\gamma|_2 = 1, \gamma \neq (0, ..., 0, 1)', \text{ and } \gamma \neq (0, ..., 0, -1)'\}$. This parameter space excludes the case of no real threshold variable by assuming that both $|\gamma|_2 = 1$ and $\gamma \neq (0, ..., 0, \pm 1)'$ (recall that the last element of f_t is -1). Assumption 2.1 is more convenient for computation since it reduces the number of unknown parameters but it requires to know which factor has a non-zero coefficient. On the other hand, the alternative parameter space might be more attractive when it is difficult to know which factor has a non-zero coefficient *a priori*. We focus on the former throughout the paper; however, the main results of the paper could be obtained under the latter.

We make the following regularity conditions.

Assumption A.1 (Identification)**.** *(i) There exists an element f_{jt} in f_t such that $\gamma_{j0} \neq 0$ and the conditional distribution of f_{jt} given $f_{-j,t}$ is continuous with probability one, where $f_{-j,t}$ is the subvector of f_t excluding f_{jt}.*

(ii) Let $B_{\gamma t} \equiv \{f_t'\gamma_0 \leq 0 < f_t'\gamma\} \cup \{f_t'\gamma \leq 0 < f_t'\gamma_0\}$. Then, for any $\gamma \in \Gamma$ such that $\gamma \neq \gamma_0$,

$$\mathbb{E}\left[\left(x_t'\delta_0\right)^2 1\{B_{\gamma t}\}\right] > 0. \tag{A.1}$$

(iii) Let $A_{1\gamma t} \equiv \{f_t'\gamma_0 > 0\} \cap \{f_t'\gamma > 0\}$ and $A_{2\gamma t} \equiv \{f_t'\gamma_0 \leq 0\} \cap \{f_t'\gamma \leq 0\}$. Then,

$$\inf_{\gamma\in\Gamma} \mathbb{E}\left[x_t x_t' 1\{A_{1\gamma t}\}\right] > 0 \quad \text{and} \quad \inf_{\gamma\in\Gamma} \mathbb{E}\left[x_t x_t' 1\{A_{2\gamma t}\}\right] > 0. \tag{A.2}$$

Recall that

$$R(\alpha, \gamma) \equiv \mathbb{E}(y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2 - \mathbb{E}(y_t - x_t'\beta_0 - x_t'\delta_0 1\{f_t'\gamma_0 > 0\})^2. \tag{A.3}$$

[1] Alternatively, it could be -1; however, the choice between $+1$ and -1 is just a labelling issue since two regimes are equivalent up to reparametrization of α_0 under either scale normalization.

Note that under Assumption A.1(i), $R(\cdot,\cdot)$ is continuous. The condition (A.1) ensures the presence of a change in the regression function. If $\delta_0 = 0$, then (A.1) is not satisfied. A sufficient condition for (A.1) is to assume that that there exists some $\eta > 0$ such that any open subset of $F_\eta \equiv \{f_t : |f_t'\gamma_0| \leq \eta\}$ possesses a positive probability (dense support) and that

$$\mathbb{E}\left[\left(x_t'\delta_0\right)^2 | f_t = z\right] > 0$$

for all but finitely many $z \in \{z : |z'\gamma_0| \leq \eta\}$ (rank condition).

The condition (A.2) is satisfied, for example, if

$$\mathbb{E}\left[x_t x_t' 1\left\{\inf_{\gamma\in\Gamma} f_t'\gamma > 0\right\}\right] > 0 \text{ and } \mathbb{E}\left[x_t x_t' 1\left\{\sup_{\gamma\in\Gamma} f_t'\gamma \leq 0\right\}\right] > 0. \tag{A.4}$$

Note that (A.4) requires that (i) the parameter space Γ satisfies

$$\mathbb{P}\left(\bigcap_{\gamma\in\Gamma}\{f_t'\gamma > 0\}\right) > 0 \text{ and } \mathbb{P}\left(\bigcap_{\gamma\in\Gamma}\{f_t'\gamma \leq 0\}\right) > 0 \tag{A.5}$$

and (ii) $\mathbb{E}\left(x_t x_t' | f_t = z\right)$ has full rank for some z belonging to $\{z : \inf_{\gamma\in\Gamma} z'\gamma > 0\}$ and also for some z such that $\left\{z : \sup_{\gamma\in\Gamma} z'\gamma \leq 0\right\}$. In other words, there should be some non-negligible fraction of observations in each regime for any $\gamma \in \Gamma$. However, we cannot simply assume that $\mathbb{E}\left(x_t x_t' | f_t = z\right) > 0$ for all z since x_t may contain f_t and thus the positive-definiteness may not hold for all z.

Remark A.2. It is possible to provide sufficient conditions for Assumption A.1 in a more compact form if x_t does not contain $f_t = (f_{1t}, f_{2t}')'$ other than the constant 1. For instance, in that case, it suffices to assume that $\delta_0 \neq 0$, the conditional distribution of f_{1t} given f_{2t} has everywhere positive density with respect to Lebesque measure for almost every f_{2t}, and both $\mathbb{E}\left(f_{2t} f_{2t}'\right) > 0$ and $\mathbb{E}\left(x_t x_t' | f_t\right) > 0$ a.s.

The following theorem gives the identification and well-separability of $(\alpha_0', \gamma_0')'$.

Theorem A.1 (Identification)**.** *If Assumptions 2.1 and A.1 hold, then* (α_0', γ_0') *is the unique solution to*

$$\min_{(\alpha',\gamma')'\in\mathbb{R}^{2d_x}\times\Gamma} \mathbb{E}(y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2$$

and

$$\inf_{\{(\alpha',\gamma')'\in\mathbb{R}^{2d_x}\times\Gamma : |(\alpha',\gamma')-(\alpha_0',\gamma_0')|_2 > \varepsilon\}} R(\alpha,\gamma) > 0$$

for any $\varepsilon > 0$.

Theorem A.1 gives the basis for our estimator given in the main text.

Proof of Theorem A.1. Note that

$$R(\alpha,\gamma) = \mathbb{E}\left(Z_t(\gamma)'\alpha - Z_t(\gamma_0)'\alpha_0\right)^2$$

due to (1.1) and (1.2). We consider two cases separately: (1) $\alpha = \alpha_0$ and $\gamma \neq \gamma_0$ and (2) $\alpha \neq \alpha_0$.

First, when $\alpha = \alpha_0$ and $\gamma \neq \gamma_0$,

$$\left(Z_t(\gamma)'\alpha - Z_t(\gamma_0)'\alpha_0\right)^2 = \left(x_t'\delta_0\right)^2$$

on $B_\gamma = \{f_t'\gamma_0 \leq 0 < f_t'\gamma\} \cup \{f_t'\gamma \leq 0 < f_t'\gamma_0\}$. Thus,

$$R(\alpha_0,\gamma) \geq \mathbb{E}\left[\left(x_t'\delta_0\right)^2 1\{B_\gamma\}\right] > 0$$

by (A.1) and $R(\alpha_0,\gamma)$ is continuous at $\gamma = \gamma_0$ due to Assumption A.1 (i).

Second, if $\alpha \neq \alpha_0$,

$$\left(Z_t(\gamma)'\alpha - Z_t(\gamma_0)'\alpha_0\right)^2 = \left(x_t'(\beta - \beta_0 + \delta - \delta_0)\right)^2$$

on $\{f_t'\gamma_0 > 0\} \cap \{f_t'\gamma > 0\}$ and

$$\left(Z_t(\gamma)'\alpha - Z_t(\gamma_0)'\alpha_0\right)^2 = \left(x_t'(\beta - \beta_0)\right)^2$$

on $\{f_t'\gamma_0 \leq 0\} \cap \{f_t'\gamma \leq 0\}$. Thus,

$$\begin{aligned} R(\alpha,\gamma) &\geq \mathbb{E}\left(x_t'(\beta - \beta_0 + \delta - \delta_0)\right)^2 1\{A_{1\gamma t}\} \\ &\quad + \mathbb{E}\left(x_t'(\beta - \beta_0)\right)^2 1\{A_{2\gamma t}\} \\ &> c\,|\alpha - \alpha_0|_2^2, \end{aligned} \tag{A.6}$$

for some $c > 0$ due to the rank condition in (A.2).

Together, they imply that the minimizer of R is unique and well-separated. ∎

B Additional Details on Computation

In this section, we provide additional details on computation. We give the proof of Theorem 2.1 and present an alternative form of the proposed algorithm in Section 2.3.

B.1 Proof for Section 2

Proof of Theorem 2.1. For convenience, we number constraints in the following way: $\forall t, j$,

1. $(\beta, \delta) \in \mathcal{A},\ \gamma \in \Gamma$,

2. $L_j \leq \delta_j \leq U_j$,

3. $(d_t - 1)(M_t + \epsilon) < f_t'\gamma \leq d_t M_t$,

4. $d_t \in \{0, 1\}$,

5. $d_t L_j \leq \ell_{j,t} \leq d_t U_j$,

6. $L_j(1 - d_t) \leq \delta_j - \ell_{j,t} \leq U_j(1 - d_t)$,

7. $\tau_1 \leq \frac{1}{T}\sum_{t=1}^T d_t \leq \tau_2$.

Recall that

$$\mathbb{Q}_T\left(\beta, \boldsymbol{\ell}\right) \equiv \frac{1}{T}\sum_{t=1}^T \left(y_t - x_t'\beta - \sum_{j=1}^{d_x} x_{j,t}\ell_{j,t}\right)^2,$$

where $\boldsymbol{\ell} = (\ell_{1,1}, \ell_{1,2}, ..., \ell_{d_x,T})'$,

$$\left(\bar{\beta}, \bar{\delta}, \bar{\gamma}, \bar{\boldsymbol{d}}, \bar{\boldsymbol{\ell}}\right) = \underset{\beta,\delta,\gamma,\boldsymbol{d},\boldsymbol{\ell}}{\operatorname{argmin}}\, \mathbb{Q}_T\left(\beta, \boldsymbol{\ell}\right) \text{ under conditions 1-7},$$

and $\mathbb{S}_T\left(\alpha, \gamma\right) \equiv \frac{1}{T}\sum_{t=1}^T (y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2$ and $\widehat{\alpha}$ and $\widehat{\gamma}$ denote the argmin of $\mathbb{S}_T$.

To prove the theorem, we show that (i) $\mathbb{S}_T\left(\bar{\alpha}, \bar{\gamma}\right) = \mathbb{Q}_T\left(\bar{\beta}, \bar{\boldsymbol{\ell}}\right)$; (ii) $\mathbb{Q}_T\left(\bar{\beta}, \bar{\boldsymbol{\ell}}\right) \geq \mathbb{S}_T\left(\widehat{\alpha}, \widehat{\gamma}\right)$; (iii) $\mathbb{S}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) \geq \mathbb{Q}_T\left(\bar{\beta}, \bar{\boldsymbol{\ell}}\right)$.

Proof of (i): By definition, $\mathbb{S}_T\left(\bar{\alpha}, \bar{\gamma}\right) = \frac{1}{T}\sum_{t=1}^T (y_t - x_t'\bar{\beta} - x_t'\bar{\delta} 1\{f_t'\bar{\gamma} > 0\})^2$. Hence we need to show

$$\frac{1}{T}\sum_{t=1}^T (y_t - x_t'\bar{\beta} - x_t'\bar{\delta} 1\{f_t'\bar{\gamma} > 0\})^2 = \frac{1}{T}\sum_{t=1}^T \left(y_t - x_t'\bar{\beta} - \sum_{j=1}^{d_x} x_{j,t}\bar{\ell}_{j,t}\right)^2.$$

We show $\bar{\ell}_{j,t} = \bar{\delta}_j 1\{f_t'\bar{\gamma} > 0\}$ for all (t, j). If $f_t'\bar{\gamma} > 0$, $\bar{d}_t = 1$ by condition 3 and 4, and $\bar{\ell}_{j,t} = \bar{\delta}_j$ by condition 6. If $f_t'\bar{\gamma} \leq 0$, $\bar{d}_t = 0$ by condtion 3 and 4 and $\bar{\ell}_{j,t} = 0$ by condtion 5.

Proof of (ii): By part (i), we have

$$\mathbb{Q}_T\left(\bar{\beta}, \bar{\boldsymbol{\ell}}\right) = \mathbb{S}_T\left(\bar{\alpha}, \bar{\gamma}\right) \geq \min_{\alpha \in \mathcal{A}, \gamma \in \Gamma} \mathbb{S}_T\left(\alpha, \gamma\right) = \mathbb{S}_T\left(\widehat{\alpha}, \widehat{\gamma}\right).$$

Proof of (iii): Define $\widehat{\ell}_{j,t} := \widehat{\delta}_j \widehat{d}_t$, where $\widehat{d}_t = 1\left\{f_t'\widehat{\gamma} > 0\right\}$. Then $\mathbb{S}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) = \mathbb{Q}_T\left(\widehat{\beta}, \widehat{\boldsymbol{\ell}}\right)$, where $\widehat{\boldsymbol{\ell}} = \left(\widehat{\ell}_{1,1}, ..., \widehat{\ell}_{d_x,T}\right)'$. Now it is straightforward to check that $\left(\widehat{\beta}, \widehat{\delta}, \widehat{\gamma}, \widehat{\boldsymbol{d}}, \widehat{\boldsymbol{\ell}}\right)$ satisfy conditions 1-7 for all j and t. For simplicity, we just give the details of checking condition 3. When $f_t'\widehat{\gamma} > 0$, then $\widehat{d}_t = 1$. Condition 3 becomes $0 < f_t'\widehat{\gamma} \leq M_t = \sup_{\gamma \in \Gamma} |f_t'\gamma|$, which is satisfied. When $f_t'\widehat{\gamma} \leq 0$, $\widehat{d}_t = 0$. Condition 3 becomes $-M_t - \epsilon < f_t'\widehat{\gamma} \leq 0$, which holds for any $\epsilon > 0$. So it is a feasible to the optimization problem $\min \mathbb{Q}_T$ with conditions 1-7. Consequently,

$$\mathbb{S}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) = \mathbb{Q}_T\left(\widehat{\beta}, \widehat{\boldsymbol{\ell}}\right) \geq \mathbb{Q}_T\left(\bar{\beta}, \bar{\boldsymbol{\ell}}\right)$$

by the definition of $(\bar{\beta}, \bar{\boldsymbol{\ell}})$. Combining parts (i),(ii) and (iii), $\mathbb{S}_T\left(\bar{\alpha}, \bar{\gamma}\right) = \mathbb{Q}_T\left(\bar{\beta}, \bar{\boldsymbol{\ell}}\right) = \mathbb{S}_T\left(\widehat{\alpha}, \widehat{\gamma}\right)$. ∎

B.2 Alternative MIQP Algorithm

The proposed algorithm in Section 2.3 may run slowly when the dimension of x_t is large. To mitigate this problem, we reformulate MIQP in the following way.

`[MIQP (Alternative Form)]` Let $\boldsymbol{d} = (d_1, \ldots, d_T)'$ and $\widetilde{\boldsymbol{\ell}} = \{\widetilde{\ell}_{j,t} : j = 1, \ldots, d_x, t = 1, \ldots, T\}$, where $\widetilde{\ell}_{j,t}$ is a real-valued variable. Solve the following problem:

$$\min_{\beta, \widetilde{\delta}, \gamma, \boldsymbol{d}, \widetilde{\boldsymbol{\ell}}} \frac{1}{T} \sum_{t=1}^{T} \left(y_t - x_t'\beta - \sum_{j=1} x_{j,t}\widetilde{\ell}_{j,t} - \left[\sum_{j=1}^{d_x} x_{j,t} L_j \right] d_t \right)^2 \tag{B.1}$$

subject to

$$\begin{aligned}
&(\beta, \delta) \in \mathcal{A}, \ \gamma \in \Gamma, \\
&0 \leq \widetilde{\delta}_j \leq (U_j - L_j), \\
&0 \leq \widetilde{\ell}_{j,t} \leq \widetilde{\delta}_j, \\
&(d_t - 1)(M_t + \epsilon) < f_t'\gamma \leq d_t M_t, \\
&d_t \in \{0, 1\}, \\
&0 \leq \sum_{j=1}^{d_x} \widetilde{\ell}_{j,t} \leq d_t \sum_{j=1}^{d_x} (U_j - L_j), \\
&0 \leq \sum_{j=1}^{d_x} \left[\widetilde{\delta}_j - \widetilde{\ell}_{j,t} \right] \leq (1 - d_t) \sum_{j=1}^{d_x} (U_j - L_j), \\
&\tau_1 \leq \frac{1}{T} \sum_{t=1}^{T} d_t \leq \tau_2
\end{aligned} \tag{B.2}$$

for each $t = 1, \ldots, T$ and each $j = 1, \ldots, d_x$, where $0 < \tau_1 < \tau_2 < 1$.

Note that $\widetilde{\delta}_j$ and $\widetilde{\ell}_{j,t}$ are transformed to be positive. Using the positivity of these variables, one can sum up restrictions across j's, where $j = 1, \ldots, d_x$, while ensuring that optimization problem (B.1) under (B.2) is mathematically equivalent to optimization problem (2.6) under

(2.7) in Section 2.3. We use the alternative form of formulation in our numerical work; however, we present a simpler form in Section 2.3 to help readers follow our basic ideas more easily.

C Selecting Relevant Factors

In this section, we consider factor selections. In applications, it is often difficult to have *a priori* knowledge regarding which variables constitute f_t in (1.1). Suppose that there are a mildly large number of factors; however, we are willing to assume that only a small number of factors are active (i.e. their γ coefficients are non-zero), although we do not know their identities. This is an unordered combinatorial selection problem, but can be easily adopted in the ℓ_0-penalization framework with the help of MIO, so long as the number of candidate factors is fixed (Bertsimas, King, and Mazumder (2016)).

To be specific, decompose $f_t = (f_{1t}', f_{2t}', -1)'$, where f_t can be either the observed factors or estimated factors [2] and $\gamma = (\gamma_1', \gamma_2', \gamma_3)'$. Assume that f_{1t} is known to be active for certainty, but f_{2t} may or may not be active. Let $p = |f_{2t}|_0$. Suppose that each element of γ_2 is bounded between known values of $\underline{\gamma_2}$ and $\overline{\gamma_2}$. Let γ_{2j} denote the j-th element of γ_2, where $j = 1, \dots, p$. Assume further that we know the lower and upper bounds, say $\underline{p}$ and $\overline{p}$, of the number of active elements of γ_2. A default choice of $(\underline{p}, \overline{p})$ is $\underline{p} = 0$ and $\overline{p} = p$; however, a strictly smaller choice of $\overline{p}$ might help estimation in practice when p is relatively large and it is plausible to assume that the maximal number of factors is much less than p.

For a given penalty parameter $\lambda > 0$ (here f_t is either observed or estimated factors), define

$$\widetilde{\gamma} = \arg\min_{\gamma \in \Gamma} \min_{\beta,\delta} \frac{1}{T} \sum_{t=1}^{T} \left(y_t - x_t'\beta - x_t'\delta 1\left\{f_t'\gamma > 0\right\}\right)^2 + \lambda |\gamma|_0 \quad \text{subject to (2.3).} \tag{C.1}$$

Computation of $\widetilde{\gamma}$ can be formulated using the following optimization.

`[MIQP with Factor Selection]` In addition to $\boldsymbol{d}$ and $\boldsymbol{\ell}$, let $\boldsymbol{e} = (e_1, \dots, e_p)'$. Choose a penalty parameter $\lambda > 0$. Then solve the following problem:

$$\min_{\beta,\delta,\gamma,\boldsymbol{d},\boldsymbol{\ell},\boldsymbol{e}} \widetilde{\mathbb{Q}}_T(\beta, \boldsymbol{\ell}) \equiv \frac{1}{T} \sum_{t=1}^{T} \left(y_t - x_t'\beta - \sum_{j=1} x_{j,t}\ell_{j,t} \right)^2 + \lambda \sum_{m=1}^{p} e_m \tag{C.2}$$

[2] For this section only, we use f_{2t} excluding -1. This is to reflect our setup where the constant term -1 is always included among active factors.

subject to (2.7) and

$$
\begin{aligned}
& e_m \underline{\gamma_2} \leq \gamma_{2m} \leq e_m \overline{\gamma_2}, \\
& \underline{p} \leq \sum_{m=1}^{p} e_m \leq \overline{p}, \\
& e_m \in \{0,1\} \quad \text{for each } m = 1, \ldots, p.
\end{aligned} \tag{C.3}
$$

Finally, re-estimate the model using only selected factors via the method given in Section 2.3.

The new indicator variable e_m turns on and off the m-th factor in estimation. The complexity of the regression model is penalized by the ℓ_0 norm ($\sum_{m=1}^{p} e_m$).

When f_t contains only observed factors, we provide selection consistency below.

Theorem C.1. *Consider the known factor case. Let $S(\gamma) = \{j : \gamma_j \neq 0\}$ and $S_0 = S(\gamma_0)$. Let Assumptions 2.1, 2.2, 3.1, and 3.2 hold. Suppose that $\lambda \to 0$, $\lambda T \to \infty$, and p is fixed. Then,*

$$\mathbb{P}\left\{S(\widetilde{\gamma}) = S_0\right\} \to 1.$$

When factors are unobservable but estimated via the PCA, the optimization (C.1) can be still used to select the estimated factors. Indeed, Bai and Ng (2008) first obtained a set of PCA factors, then applied BIC to select among them for diffusion index forecasts. In theory, however, interpretation of each estimated factor is more involved since factors are identified only up to some random rotation. The rotation indeterminacy creates difficulties to define "the true factors" statistically, and therefore the "factor selection consistency". Thus we do not pursue the selection consistency in the estimated factor case.

Proof of Theorem C.1. For a given γ, let

$$\mathbb{Q}_T(\gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} \left(y_t - x_t' \widehat{\beta}(\gamma) - x_t' \widehat{\delta}(\gamma) 1\left\{f_t'\gamma > 0\right\} \right)^2$$

and

$$\widetilde{\mathbb{Q}}_T(\gamma) = \mathbb{Q}_T(\gamma) + \lambda |\gamma|_0 ,$$

where $\widehat{\alpha}(\gamma) = \left(\widehat{\beta}(\gamma)', \widehat{\delta}(\gamma)'\right)'$ is the OLS estimate of α for the given γ. The former is a profiled criterion function of the original criterion. Define

$$\widetilde{\gamma} = \arg\min_{\gamma} \widetilde{\mathbb{Q}}_T(\gamma).$$

Our proof is divided into the following steps.

Step 1. Show that $S_0 \subset S(\widetilde{\gamma})$ with probability approaching one.

Step 2. Show that $\min_{\gamma:S(\gamma)=S_0} \mathbb{Q}_T(\gamma) \leq \min_\gamma \mathbb{Q}_T(\gamma) + O_P(T^{-1})$.

Step 3. Show that for $\Gamma_b := \{\gamma : S_0 \subset S(\gamma), S_0 \neq S(\gamma)\}$,

$$\min_{\gamma\in\Gamma_b} \widetilde{\mathbb{Q}}_T(\gamma) - \min_{\gamma:S(\gamma)=S_0} \widetilde{\mathbb{Q}}_T(\gamma) > \lambda/2$$

with probability approaching one.

Now suppose $S_0 \neq S(\widetilde{\gamma})$. Then by step 1, $\widetilde{\gamma} \in \Gamma_b$, then by step 3,

$$\widetilde{\mathbb{Q}}_T(\widetilde{\gamma}) \geq \min_{\gamma\in\Gamma_b} \widetilde{\mathbb{Q}}_T(\gamma) > \min_{\gamma:S(\gamma)=S_0} \widetilde{\mathbb{Q}}_T(\gamma) + \lambda/2,$$

which contradicts with the definition of $\widetilde{\gamma}$. Consequently, we must have $S_0 = S(\widetilde{\gamma})$ with probability approaching one. ∎

Proof of Step 1. Let $\alpha^*(\gamma) = \left(\mathbb{E}Z_t(\gamma) Z_t(\gamma)'\right)^{-1} \mathbb{E}Z_t(\gamma) Z_t(\gamma_0)' \alpha_0$. Also let

$$\mathbb{Q}(\gamma) \equiv \mathbb{E}\left(y_t - Z_t(\gamma)' \alpha^*(\gamma)\right)^2 = \sigma^2 + \mathbb{E}\left(\alpha^*(\gamma)' Z_t(\gamma) - \alpha_0' Z_t(\gamma_0)\right)^2.$$

Then, by the ULLN and the CMT and the fact that $\lambda \to 0$, uniformly in γ,

$$\widehat{\alpha}(\gamma) - \alpha^*(\gamma) = o_P(1), \qquad \widetilde{\mathbb{Q}}_T(\gamma) - \mathbb{Q}(\gamma) = o_P(1).$$

Also, $\alpha^*(\gamma_0) = \alpha_0$ implies $\mathbb{Q}(\gamma_0) = \sigma^2$ and

$$\mathbb{Q}(\widetilde{\gamma}) = \widetilde{\mathbb{Q}}_T(\widetilde{\gamma}) + o_P(1) \leq \widetilde{\mathbb{Q}}_T(\gamma_0) + o_P(1) = \mathbb{Q}(\gamma_0) + o_P(1) = \sigma^2 + o_P(1).$$

On the other hand, for $\Gamma_a = \{\gamma : S_0 \nsubseteq S(\gamma)\}$, due to Theorem A.1,

$$\min_{\gamma\in\Gamma_a} \mathbb{E}\left(\alpha^*(\gamma)' Z_t(\gamma) - \alpha_0' Z_t(\gamma_0)\right)^2 > 0.$$

So $\min_{\gamma\in\Gamma_a} \mathbb{Q}(\gamma) > \sigma^2$. This implies $\widetilde{\gamma} \notin \Gamma_a$, thus $S_0 \subset S(\widetilde{\gamma})$ with probability approaching one. ∎

Proof of Step 2. Uniformly over pairs (γ_1, γ_2) in a shrinking neighborhood of γ_0, ($B_C(\gamma_0) = \{|\gamma - \gamma_0|_2 \leq CT^{-(1-2\varphi)}\}$ for any $C > 0$),

$$\mathbb{Q}_T(\gamma_1) - \mathbb{Q}_T(\gamma_2) = R_T(\gamma_1) - R_T(\gamma_2) + \mathbb{G}_T(\gamma_2) - \mathbb{G}_T(\gamma_1),$$

where $R_T(\gamma) = \frac{1}{T}\sum_t [Z_t(\gamma)'\widehat{\alpha}(\gamma) - Z_t(\gamma_0)'\alpha_0]^2$ and $\mathbb{G}_T(\gamma) = \frac{2}{T}\sum_t \varepsilon_t Z_t(\gamma)\widehat{\alpha}(\gamma)$. Note that $\sup_{\gamma\in B_C(\gamma_0)} |\widehat{\alpha}(\gamma) - \alpha_0|_2 = O_P(T^{-1/2})$, $\sup_{\gamma\in B_C(\gamma_0)} |R_T(\gamma)| = O_P(T^{-1})$, and
$\sup_{\gamma_1,\gamma_2\in B_C(\gamma_0)} |\mathbb{G}_T(\gamma_1) - \mathbb{G}_T(\gamma_2)| = O_P(T^{-1})$. Therefore,

$$\sup_{\gamma_1,\gamma_2\in B_C(\gamma_0)} |\mathbb{Q}_T(\gamma_1) - \mathbb{Q}_T(\gamma_2)| = O_P(T^{-1}).$$

Let $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ respectively denote the argument of $\min_{S(\gamma)=S_0} \mathbb{Q}_T(\gamma)$ and $\min_\gamma \mathbb{Q}_T(\gamma)$. Then for both $j = 1, 2$, $\mathbb{Q}_T(\widehat{\gamma}_j) \leq \mathbb{Q}_T(\gamma_0)$. Then it follows from the proof of Theorem 3.1 that $\widehat{\gamma}_j - \gamma_0 = O_P(T^{-(1-2\varphi)})$, $\quad j = 1, 2$. As a result,

$$0 \leq \min_{\gamma:S(\gamma)=S_0} \mathbb{Q}_T(\gamma) - \min_\gamma \mathbb{Q}_T(\gamma) = \mathbb{Q}_T(\widehat{\gamma}_1) - \mathbb{Q}_T(\widehat{\gamma}_2) = O_P(T^{-1}).$$

∎

Proof of Step 3. Let $\Gamma_b := \{\gamma : S_0 \subset S(\gamma), S_0 \neq S(\gamma)\}$. Then we have

$$\begin{aligned}
\min_{\gamma\in\Gamma_b} \widetilde{\mathbb{Q}}_T(\gamma) - \min_{\gamma:S(\gamma)=S_0} \widetilde{\mathbb{Q}}_T(\gamma) &\geq^{(1)} \min_\gamma \mathbb{Q}_T(\gamma) + \lambda \min_{\gamma\in\Gamma_b} |\gamma|_0 - \min_{\gamma:S(\gamma)=S_0} \widetilde{\mathbb{Q}}_T(\gamma) \\
&=^{(2)} \min_\gamma \mathbb{Q}_T(\gamma) - \min_{S(\gamma)=S_0} \mathbb{Q}_T(\gamma) + \lambda \min_{\gamma\in\Gamma_b} |\gamma|_0 - \lambda|\gamma_0|_0 \\
&\geq^{(3)} O_P(T^{-1}) + \lambda \\
&>^{(4)} \lambda/2 \quad \text{(with probability approaching one)}
\end{aligned}$$

where (1) is due to $\min_{\gamma\in\Gamma_b} \widetilde{\mathbb{Q}}_T(\gamma) \geq \min_\gamma \mathbb{Q}_T(\gamma) + \lambda \min_{\gamma\in\Gamma_b} |\gamma|_0$; (2) is due to the fact that $\arg\min_{\gamma:S(\gamma)=S_0} \widetilde{\mathbb{Q}}_T(\gamma) = \arg\min_{\gamma:S(\gamma)=S_0} \mathbb{Q}_T(\gamma)$, and $|\gamma|_0 = |\gamma_0|_0$ for all $\gamma \in \{\gamma : S(\gamma) = S_0\}$; (3) is due to step 2 and $\min_{\gamma\in\Gamma_b} |\gamma|_0 - |\gamma_0|_0 \geq 1$. Finally, (4) is due to $T\lambda \to \infty$. ∎

D Test of Linearity

In some applications, we are interested in testing the linearity of the regression model in (1.1). That is, we may want to test the following null hypothesis:

$$\mathcal{H}_0 : \delta_0 = 0 \quad \text{for all } \gamma_0 \in \Gamma.$$

Under the null hypothesis the model becomes the linear regression model and thus γ_0 is not identified. This testing problem has been studied intensively in the literature when f_t is directly observed and the dimension of an unidentifiable component of γ_0 is 1 (see, e.g., Hansen (1996) and Lee, Seo, and Shin (2011) among many others).

We propose to use the following statistic:

$$\begin{aligned}\text{supQ} &= \sup_{\gamma\in\Gamma} T\frac{\min_{\alpha:\delta=0}\mathbb{S}_T(\alpha,\gamma)-\min_{\alpha}\mathbb{S}_T(\alpha,\gamma)}{\min_{\alpha}\mathbb{S}_T(\alpha,\gamma)} \\ &= T\frac{\min_{\alpha:\delta=0}\mathbb{S}_T(\alpha,\gamma)-\min_{\alpha,\gamma}\mathbb{S}_T(\alpha,\gamma)}{\min_{\alpha,\gamma}\mathbb{S}_T(\alpha,\gamma)},\end{aligned} \tag{D.1}$$

where $\mathbb{S}_T(\alpha,\gamma)$ is the least squares criterion function, using either the observed or estimated factor.

For both observed and latent factor cases, we establish the following result.

Theorem D.1. *Suppose that either assumptions of Theorems 3.1 (for the known factor case) or assumptions of Theorems 4.3 (for the estimated factor case) hold. Then, under $\mathcal{H}_0$,*

$$supQ \xrightarrow{d} \sup_{\gamma\in\Gamma} W(\gamma)'\left(R\left(\mathbb{E}Z_t(\gamma)Z_t(\gamma)'\right)^{-1}\mathbb{E}\varepsilon_t^2R'\right)^{-1}W(\gamma),$$

where $W(\gamma)$ is a vector of centered Gaussian processes with covariance kernel

$$K(\gamma_1,\gamma_2) = R\left(\mathbb{E}Z_t(\gamma_1)Z_t(\gamma_1)'\right)^{-1}\mathbb{E}\left[Z_t(\gamma_1)Z_t(\gamma_2)'\varepsilon_t^2\right]\left(\mathbb{E}Z_t(\gamma_2)Z_t(\gamma_2)'\right)^{-1}R'$$

and $R=(0_{d_x},I_{d_x})$ is the $(d_x\times 2d_x)$-dimensional selection matrix.[3]

Below we present a bootstrap algorithm for the p-value.

`[Computation of Bootstrap` p`-Values]`

1. Generate an iid sequence $\{\eta_t\}$ whose mean is zero and variance is one.

2. Construct $\{y_t^*\}$ by
$$y_t^* = x_t'\widehat{\beta}+\eta_t\widehat{\varepsilon}_t,$$
where $\widehat{\beta}$ is the unconstrained estimator of β_0 and $\widehat{\varepsilon}_t$ is the estimated residual from unconstrained estimation.

3. Construct the bootstrap statistic supQ* by (D.1) with the bootstrap sample $\{y_t^*,x_t,f_t : t=1,\ldots,T\}$ if f_t is known and $\{y_t^*,x_t,\widetilde{f}_t : t=1,\ldots,T\}$ if f_t is estimated, respectively.

4. Repeat 1-3 many times and compute the empirical distribution of supQ*.

5. Then, with the obtained empirical distribution, say $F_T^*(\cdot)$, one can compute the bootstrap p-value by
$$p^* = 1-F_T^*(\text{supQ}),$$

[3]Here, 0_{d_x} and I_{d_x}, respectively, denote the d_x-dimensional square matrix with all elements being zeros and the d_x-dimensional identity matrix.

or a-level critical value

$$c_a^* = F_T^{*^{-1}} (1 - a) .$$

The proposed bootstrap is standard and thus its asymptotic validity follows from the standard manner in view of Lemma G.1 and the conditional martingale difference sequence central limit theorem (e.g. Theorem 3.2 of Hall and Heyde (1980)). The details are omitted for the sake of brevity. Furthermore, it is straightforward to establish conditions for the consistency of our proposed test.

E Additional Empirical Results

In this part of the appendix, we provide additional empirical results that are omitted from the main text.

E.1 Testing the Linearity of US GNP and Selecting Factors

In this section, we revisit the empirical application in Hansen (1996), who tested Potter (1995)'s model of US GNP. Hansen (1996) used annualized quarterly growth rates, say y_t, for the period 1947-1990. His estimates were as follows:

$$\begin{aligned}
y_t &= \underset{(2.12)}{-3.21} + \underset{(0.25)}{0.51} y_{t-1} - \underset{(0.31)}{0.93} y_{t-2} - \underset{(0.25)}{0.38} y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} \leq 0.01 \\
y_t &= \underset{(0.77)}{2.14} + \underset{(0.10)}{0.30} y_{t-1} + \underset{(0.10)}{0.18} y_{t-2} - \underset{(0.07)}{0.16} y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} > 0.01,
\end{aligned} \tag{E.1}$$

where heteroskedasticity-robust standard errors are given in parenthesis. His heteroskedasticity-robust LM-based tests for the hypothesis of no threshold effect were all far from usual rejection regions (the smallest p-value was 0.17). Using the same dataset, we carry out the following two exercises: (1) selecting relevant factors and (2) testing the linearity of the model. For the former, we keep y_{t-2} as f_{1t} and add (y_{t-1}, y_{t-5}) as f_{2t}. That is, we allow for the possibility that the regimes can be determined by a linear combination of $(y_{t-1}, y_{t-2}, y_{t-5})$. The choice of penalization parameter λ is important. Recall that we require $\lambda \to 0$ and $\lambda T \to \infty$. In this application, we set

$$\lambda = \widehat{\sigma}^2_{\text{Hansen}} \frac{\log T}{T},$$

where $\widehat{\sigma}^2_{\text{Hansen}} = T^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2$ and the estimated residual $\widehat{\varepsilon}_t$ is obtained from Hansen (1996)'s estimates in (E.1). By implementing joint optimization with this choice of λ, we select only y_{t-5} but drop y_{t-1} in f_{2t}. Our estimated index is

$$f_t'\widehat{\gamma} = y_{t-2} - 0.91y_{t-5} + 0.50.$$

If we compare this with Hansen's estimate $f_t'\widehat{\gamma} = y_{t-2} - 0.01$, we can see that in Hansen's model, the regime is determined by the level of GNP growth in $t-2$; on the contrary, in our model, it is determined by $y_{t-2} - 0.91y_{t-5}$, roughly speaking the changes in growth rates from $t-5$ to $t-2$. Specifically, the regime is determined whether $y_{t-2} - 0.91y_{t-5}$ is above or below -0.50. Our estimates suggest that a recession might be captured better by a decrease in growth rates from $t-5$ to $t-2$, compared to a low level of growth rates in $t-2$. Our estimated coefficients and their standard errors are as follows:

$$\begin{aligned} y_t &= \underset{(1.33)}{-2.07} + \underset{(0.13)}{0.28}y_{t-1} - \underset{(0.16)}{0.33}y_{t-2} + \underset{(0.19)}{0.62}y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} - 0.91y_{t-5} \leq -0.50 \\ y_t &= \underset{(0.96)}{2.76} + \underset{(0.12)}{0.35}y_{t-1} + \underset{(0.12)}{0.07}y_{t-2} - \underset{(0.10)}{0.21}y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} - 0.91y_{t-5} > -0.50. \end{aligned} \tag{E.2}$$

We now report the result of testing the null hypothesis of no threshold effect. We take our estimates in (E.2) as unconstrained estimates. The resulting LR test statistic is 28.19 and the p-value is 0.056 based on 500 bootstrap replications. This implies that the null hypothesis is rejected at the 10% level but not at the 5% level. There are two main differences between our test result and Hansen (1996)'s. We use the LR statistic, whereas Hansen (1996) considered the LM statistic. Furthermore, his alternative only allows for the scalar threshold variable y_{t-2} but we consider a single index using y_{t-2} and y_{t-5}.

E.2 Details on Estimation Results for Table 2

Table A-1 shows full estimation results for Table 2.

F Proofs of the Asymptotic Distribution in Section 3: Known f

Recall that we have proposed two computing algorithms for the estimators of (α_0, γ_0). Throughout this part of the appendix, we assume that $(\widehat{\alpha}, \widehat{\gamma})$ is the global solution to the optimization problem. The proof is divided into the following subsections.

Table A-1: Estimation Results

Specification	(1)		(2)		(3)	
	$f_{1t} = (q_{t-1}, -1)$		$f_{2t} = (F_{t-1}, -1)$		$f_{3t} = (q_{t-1}, F_{t-1}, -1)$	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Regime 1 ("Expansion")	$q_{t-1} \leq 0.302$		$F_{t-1} \leq -0.28$		$q_{t-1} + 3.55F_{t-1} \leq -1.60$	
Intercept	-0.0214	0.0126	-0.0255	0.0101	-0.0294	0.0101
Δu_{t-1}	-0.1696	0.0640	-0.1182	0.0629	-0.1628	0.0601
Δu_{t-2}	0.0382	0.0650	0.0774	0.0558	0.0264	0.0600
Δu_{t-3}	0.1896	0.0587	0.2097	0.0645	0.1933	0.0520
Δu_{t-4}	0.1399	0.0630	0.1039	0.0523	0.1445	0.0552
Δu_{t-5}	0.0858	0.0749	0.0622	0.0600	0.0699	0.0656
Δu_{t-6}	0.0214	0.0653	0.0193	0.0558	0.0177	0.0613
Δu_{t-7}	0.0318	0.0678	-0.0268	0.0596	0.0174	0.0613
Δu_{t-8}	0.0402	0.0599	-0.0006	0.0617	0.0103	0.0626
Δu_{t-9}	-0.0667	0.0663	-0.0766	0.0660	-0.0637	0.0656
Δu_{t-10}	-0.0540	0.0640	-0.0120	0.0559	-0.0467	0.0575
Δu_{t-11}	0.0782	0.0568	0.0162	0.0529	0.0196	0.0528
Δu_{t-12}	-0.0899	0.0641	-0.1216	0.0576	-0.1224	0.0572
Regime 2 ("Contraction")	$q_{t-1} > 0.302$		$F_{t-1} > -0.28$		$q_{t-1} + 3.55F_{t-1} > -1.60$	
Intercept	0.0876	0.0375	0.0509	0.0560	0.1893	0.0576
Δu_{t-1}	0.2406	0.1179	0.3671	0.2011	0.2937	0.1665
Δu_{t-2}	0.2455	0.0932	0.2198	0.1634	0.1420	0.1279
Δu_{t-3}	0.1283	0.1038	0.0936	0.1563	0.1042	0.1549
Δu_{t-4}	-0.0222	0.1033	-0.0053	0.1883	-0.1035	0.1690
Δu_{t-5}	-0.0272	0.1104	-0.1804	0.2188	-0.0723	0.1868
Δu_{t-6}	-0.0851	0.1083	-0.0500	0.2125	-0.0821	0.1400
Δu_{t-7}	-0.1562	0.1057	-0.0297	0.2027	-0.1853	0.1443
Δu_{t-8}	-0.0372	0.1357	0.0021	0.2923	-0.1214	0.2038
Δu_{t-9}	0.0991	0.1358	0.0754	0.1754	-0.0861	0.1475
Δu_{t-10}	0.1149	0.1125	0.0445	0.1574	0.0392	0.1426
Δu_{t-11}	-0.1012	0.1256	0.1872	0.1995	-0.0307	0.1840
Δu_{t-12}	-0.4440	0.1144	-0.2269	0.1668	-0.3807	0.1542
Avg. of squared residuals ($T^{-1}\sum_{i=1}^{T}\hat{\varepsilon}_t^2$)	0.0264		0.0272		0.0252	
Proportion of matches between NBER recession dates and threshold estimates	0.807		0.894		0.896	

F.1 Consistency

Lemma F.1 (Consistency)**.** *Let Assumptions 2.1, A.1 and 3.1 (i) and (ii) hold. Then as* $T \to \infty$*,*

$$|\widehat{\alpha} - \alpha_0|_2 = o_P(1) \text{ and } |\widehat{\gamma} - \gamma_0|_2 = o_P(1).$$

Proof of Lemma F.1. We begin with stating the following standard ULLN for ρ-mixing sequences, see e.g. Davidson (1994), for which Assumption 3.1 (i) and (ii) suffice.

(i) $\sup_{\gamma \in \Gamma} |\frac{1}{T} \sum_{t=1}^{T} Z_{ti}(\gamma) Z_{tj}(\gamma) - \mathbb{E}[Z_{ti}(\gamma) Z_{tj}(\gamma)]| = o_P(1)$.

(ii) $\sup_{\gamma \in \Gamma} |\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t Z_t(\gamma)| = o_P(1)$.

These will be cited as ULLN hereafter.

We begin with the consistency of $\widehat{\gamma}$. Recall that the least squares estimate of α for a given γ is the OLS estimate and construct the profiled least squares criterion $\mathbb{S}_T(\gamma)$, that is,

$$\begin{aligned} \mathbb{S}_T(\gamma) &= \mathbb{S}_T(\widehat{\alpha}(\gamma), \gamma) = \frac{1}{T} Y'(I - P(\gamma)) Y \\ &= \frac{1}{T} \left(e'(I - P(\gamma)) e + 2\delta_0' X_0 (I - P(\gamma)) e + \delta_0' X_0' (I - P(\gamma)) X_0 \delta_0 \right), \end{aligned}$$

where e, Y, and X_0 are the matrices stacking ε_t's, y_t's and $x_t' 1_t$'s, respectively, and $P(\gamma)$ is the orthogonal projection matrix onto $Z_t(\gamma)$'s.

Let $\widetilde{\gamma}$ be an estimator such that

$$\mathbb{S}_T(\widetilde{\gamma}) \leq \mathbb{S}_T(\gamma_0) + o_P\left(T^{-2\varphi}\right). \tag{F.1}$$

Then, by Lemma F.2, the ULLN for $T^{-1} \sum_{t=1}^{T} Z_t(\gamma) Z_t(\gamma)'$, the rank condition for $\mathbb{E} Z_t(\gamma) Z_t(\gamma)'$ in Assumption 3.1 (iii), the fact that $P(\gamma_0) X_0 = X_0$,

$$\begin{aligned} 0 &\geq T^{2\varphi} (\mathbb{S}_T(\widetilde{\gamma}) - \mathbb{S}_T(\gamma_0)) - o_P(1) \\ &= \frac{T^{2\varphi}}{T} \left(e'(P(\gamma_0) - P(\widetilde{\gamma})) e + 2\delta_0' X_0 (P(\gamma_0) - P(\widetilde{\gamma})) e + \delta_0' X_0' (P(\gamma_0) - P(\widetilde{\gamma})) X_0 \delta_0 \right) \\ &= o_P(1) + \frac{1}{T} d_0' X_0' (I - P(\widetilde{\gamma})) X_0 d_0, \\ &= o_P(1) + \mathbb{E} d_0' x_t x_t' d_0 1_t - \underbrace{\left(\mathbb{E} d_0' x_t 1_t Z_t(\widetilde{\gamma})' \right) \left(\mathbb{E} Z_t(\widetilde{\gamma}) Z_t(\widetilde{\gamma})' \right)^{-1} \mathbb{E} Z_t(\widetilde{\gamma}) 1_t x_t' d_0}_{A(\widetilde{\gamma})}. \end{aligned}$$

However, the term $A(\widetilde{\gamma})$ is continuous by Assumption A.1 and has maximum at $\widetilde{\gamma} = \gamma_0$ by the property of the orthogonal projection, and $\mathbb{E} d_0' x_t x_t' d_0 1_t - A(\gamma) > 0$ for any $\gamma \neq \gamma_0$ due to Assumptions A.1 (ii) and 3.1 (iii). Finally, the compact parameter space yields the consistency of $\widehat{\gamma}$ by the argmax continuous mapping theorem (see, e.g., van der Vaart and Wellner (1996, p.286)).

Turning to $\widehat{\alpha}$, note that

$$\begin{aligned} 0 &\geq \mathbb{S}_T(\widehat{\alpha}, \widehat{\gamma}) - \mathbb{S}_T(\alpha_0, \gamma_0) \\ &= \mathbb{R}_T(\widehat{\alpha}, \widehat{\gamma}) - \mathbb{G}_T(\widehat{\alpha}, \widehat{\gamma}) + \mathbb{G}_T(\alpha_0, \gamma_0), \end{aligned} \tag{F.2}$$

where

$$\mathbb{R}_T(\alpha, \gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} \left(Z_t(\gamma)' \alpha - Z_t(\gamma_0)' \alpha_0 \right)^2$$
$$\mathbb{G}_T(\alpha, \gamma) \equiv \frac{2}{T} \sum_{t=1}^{T} \varepsilon_t Z_t(\gamma)' \alpha.$$

First, note that

$$\begin{aligned} &\mathbb{R}_T(\alpha, \gamma) - R(\alpha, \gamma) \\ &= (\alpha - \alpha_0)' \frac{1}{T} \sum_{t=1}^{T} \left(Z_t(\gamma) Z_t(\gamma)' - \mathbb{E} Z_t(\gamma) Z_t(\gamma)' \right) (\alpha - \alpha_0) \\ &+ \frac{1}{T} \sum_{t=1}^{T} \left(x_t' \delta_0 \right)^2 |1_t(\gamma) - 1_t(\gamma_0)| - \mathbb{E} \left(x_t' \delta_0 \right)^2 |1_t(\gamma) - 1_t(\gamma_0)| \\ &+ \frac{2\delta_0'}{T} \sum_{t=1}^{T} \Big[x_t (1_t(\gamma) - 1_t(\gamma_0)) Z_t(\gamma) - \mathbb{E}[x_t (1_t(\gamma) - 1_t(\gamma_0)) Z_t(\gamma)] \Big]' (\alpha - \alpha_0) \\ &= o_P(1)(|\alpha - \alpha_0|_2^2 + |\alpha - \alpha_0|_2) \text{ uniformly in } \gamma \in \Gamma, \end{aligned} \tag{F.3}$$

by ULLN. Similarly,

$$\begin{aligned} &\mathbb{G}_T(\alpha, \gamma) - \mathbb{G}_T(\alpha_0, \gamma_0) \\ &= \frac{2}{T} \sum_{t=1}^{T} \varepsilon_t Z_t(\gamma)' (\alpha - \alpha_0) + \frac{2}{T} \sum_{t=1}^{T} \varepsilon_t x_t' \delta_0 (1_t(\gamma) - 1_t(\gamma_0)), \\ &= o_P(1)(|\alpha - \alpha_0|_2) \text{ uniformly in } \gamma \in \Gamma \end{aligned} \tag{F.4}$$

Combining these results together implies that

$$R(\widehat{\alpha}, \widehat{\gamma}) \leq o_P(1)(|\widehat{\alpha} - \alpha_0|_2 + |\widehat{\alpha} - \alpha_0|_2^2).$$

Then, combining this result with the proof of Theorem A.1 implies that $\widehat{\alpha} - \alpha_0 = o_P(1)$ as (A.6) shows that R is bounded below by some positive constant times $|\alpha - \alpha_0|_2^2$. ∎

F.2 Rates of Convergence

To begin with, we assume γ belongs to a small neighborhood of γ_0 due to the preceding consistency proof. It is useful to introduce additional notation. Let $1_t(\gamma) \equiv 1\{f_t'\gamma > 0\}$ while $1_t \equiv 1_t(\gamma_0)$. Similarly, let $1_t(\gamma, \bar{\gamma}) \equiv 1\{f_t'\gamma \leq 0 < f_t'\bar{\gamma}\}$. Clearly, $1_t(\gamma) = 1_t(0, \gamma)$.

Define

$$H_{1,t}(\gamma) := \varepsilon_t x_t'\delta_0 \left(1_t(\gamma) - 1_t(\gamma_0)\right),$$
$$H_{2,t}(\gamma) := \left(x_t'\delta_0\right)^2 \left|1_t(\gamma) - 1_t(\gamma_0)\right|,$$
$$H_{3,t}(\gamma) := \left(x_t'\delta_0\right)\left(1_t(\gamma) - 1_t(\gamma_0)\right) Z_{tj}(\gamma),$$

where $Z_{tj}(\gamma)$ is the j-th element of $Z_t(\gamma)$. For the simplicity of notation, we suppress the dependence of $H_{3,t}(\gamma)$ on j. We first state a lemma that is a direct consequence of Lemmas I.1 and I.2 for an easy reference.

Lemma F.2. *There exists a constant $C_2 > 0$ such that for any $\eta > 0$,*

$$\sup_{|\gamma-\gamma_0|_2 \leq T^{-1+2\varphi}} \left| \frac{1}{T}\sum_{t=1}^{T} \left\{H_{k,t}(\gamma) - \mathbb{E}H_{k,t}(\gamma)\right\} \right| = O_P\left(\frac{1}{T}\right),$$
$$\sup_{|\gamma-\gamma_0|_2 \leq T^{-1+2\varphi}} \left| \frac{1}{T}\sum_{t=1}^{T} \left\{H_{2,t}(\gamma) - \mathbb{E}H_{2,t}(\gamma)\right\} \right| = O_P\left(\frac{1}{T^{1+\varphi}}\right),$$
$$\sup_{T^{-1+2\varphi} < |\gamma-\gamma_0|_2 < C_2} \left| \left| \frac{1}{T}\sum_{t=1}^{T} \left\{H_{k,t}(\gamma) - \mathbb{E}H_{k,t}(\gamma)\right\} \right| - \eta T^{-2\varphi} |\gamma-\gamma_0|_2 \right| = O_P\left(\frac{1}{T}\right),$$

where $k = 1, 2, 3$.

Lemma F.3 (Rates of Convergence)**.** *Let Assumptions 2.1, A.1, 3.1, and 3.2 hold. Then as $T \to \infty$,*

$$|\widehat{\alpha} - \alpha_0|_2 = O_P\left(\frac{1}{\sqrt{T}}\right) \text{ and } |\widehat{\gamma} - \gamma_0|_2 = O_P\left(\frac{1}{T^{1-2\varphi}}\right).$$

Proof of Lemma F.3. The proof is based on the following two steps, which will be shown later.

Step 1. As $T \to \infty$, there exist positive constants c and e, with probability approaching one,

$$R(\alpha, \gamma) \geq c\,|\alpha - \alpha_0|_2^2 + cT^{-2\varphi}\,|\gamma - \gamma_0|_2,$$

for any α and γ such that $|\alpha - \alpha_0| < e$ and $|\gamma - \gamma_0| < e$. Recall $R(\alpha, \gamma)$ is defined in (A.3).

Step 2. There exists a positive constant $\eta < c/2$ such that

$$|\mathbb{G}_T(\alpha,\gamma) - \mathbb{G}_T(\alpha_0,\gamma_0)| \le O_P\left(\frac{1}{\sqrt{T}}\right)|\alpha-\alpha_0|_2 + \eta T^{-2\varphi}|\gamma-\gamma_0|_2 + O_P\left(\frac{1}{T}\right) \tag{F.5}$$

$$|\mathbb{R}_T(\alpha,\gamma) - R(\alpha,\gamma)| \le \eta|\alpha-\alpha_0|_2^2 + \eta T^{-2\varphi}|\gamma-\gamma_0|_2 + O_P\left(\frac{1}{T}\right), \tag{F.6}$$

where the inequalities above are uniform in α and γ such that $|\alpha-\alpha_0| < e$ and $|\gamma-\gamma_0| < e$, in the sense that the sequences $O_P(\cdot)$ and $o_P(\cdot)$ do not depend on α and γ.

Given Steps 1 and 2, since

$$R(\widehat{\alpha},\widehat{\gamma}) \le |\mathbb{G}_T(\widehat{\alpha},\widehat{\gamma}) - \mathbb{G}_T(\alpha_0,\gamma_0)| + |\mathbb{R}_T(\widehat{\alpha},\widehat{\gamma}) - R(\widehat{\alpha},\widehat{\gamma})|,$$

we conclude that

$$(c-2\eta)\left(|\widehat{\alpha}-\alpha_0|_2^2 + T^{-2\varphi}|\widehat{\gamma}-\gamma_0|_2\right) \le O_P\left(\frac{1}{\sqrt{T}}\right)|\widehat{\alpha}-\alpha_0|_2 + O_P\left(\frac{1}{T}\right). \tag{F.7}$$

That is,

$$|\widehat{\alpha}-\alpha_0|_2^2 \le O_P\left(\frac{1}{\sqrt{T}}\right)|\widehat{\alpha}-\alpha_0|_2 + O_P\left(\frac{1}{T}\right),$$

implying

$$|\widehat{\alpha}-\alpha_0|_2 = O_P\left(\frac{1}{\sqrt{T}}\right) \text{ and thus } |\widehat{\gamma}-\gamma_0|_2 = O_P\left(\frac{1}{T^{1-2\varphi}}\right).$$

∎

Proof of Step 1. Due to Assumption 3.2 and then Assumption A.1 we can find positive constants c, c_0 such that

$$\begin{aligned}\mathbb{E}\left(x_t'\delta_0\left(1_t(\gamma)-1_t(\gamma_0)\right)\right)^2 &\ge T^{-2\varphi}c\mathbb{E}\,|1_t(\gamma)-1_t(\gamma_0)| \\ &\ge c_0T^{-2\varphi}|\gamma-\gamma_0|_2.\end{aligned}$$

More specifically, we need to show that there exists a constant $c > 0$ and a neighborhood of γ_0 such that for all γ in the neighborhood

$$G(\gamma) = \mathbb{E}\,|1_t(\gamma)-1_t(\gamma_0)| \ge c\,|\gamma-\gamma_0|_2.$$

Note that $f_t'\gamma_0 = u_t$ and the first element of $(\gamma-\gamma_0)$ is zero due to the normalization. Then,

$$G(\gamma) = \mathbb{P}\left\{-f_{2t}'(\gamma_2-\gamma_{20}) \le u_t < 0\right\} + \mathbb{P}\left\{0 < u_t \le -f_{2t}'(\gamma_2-\gamma_{20})\right\}.$$

Since the conditional density of u_t is bounded away from zero and continuous, we can find a strictly positive lower bound, say c_1, of the conditional density of u_t if we choose a sufficiently small open neighborhood ϵ of zero. Then,

$$\mathbb{P}\left\{-f_{2t}'\left(\gamma_2-\gamma_{20}\right)\leq u_t<0\right\}\geq c_1\mathbb{E}\left(f_{2t}'\left(\gamma_2-\gamma_{20}\right)1\left\{f_{2t}'\left(\gamma_2-\gamma_{20}\right)>0\right\}1\left\{\left|f_{2t}'\right|\leq M\right\}\right),$$

where M satisfies that $\max\left|\gamma-\gamma_0\right|_2 M$ belongs to ϵ. This is always feasible because we can make $\max\left|\gamma-\gamma_0\right|_2$ as small as necessary due to the consistency of $\hat{\gamma}$. Similarly,

$$\mathbb{P}\left\{0<u_t\leq-f_{2t}'\left(\gamma_2-\gamma_{20}\right)\right\}\geq c_1\mathbb{E}\left(-f_{2t}'\left(\gamma_2-\gamma_{20}\right)1\left\{f_{2t}'\left(\gamma_2-\gamma_{20}\right)<0\right\}1\left\{\left|f_{2t}'\right|\leq M\right\}\right).$$

Thus,

$$G\left(\gamma\right)\geq c_1\mathbb{E}\left(\left|f_{2t}'\left(\gamma_2-\gamma_{20}\right)\right|1\left\{\left|f_{2t}'\right|\leq M\right\}\right)\geq c_2\left|\gamma-\gamma_0\right|_2$$

for some $c_2>0$ because

$$\inf_{|r|=1}\mathbb{E}\left(\left|f_{2t}'r\right|1\left\{\left|f_{2t}'\right|\leq M\right\}\right)>0$$

for some $M<\infty$ due to Assumption 3.2.

Next,

$$\mathbb{E}\left(Z_t\left(\gamma\right)'\left(\alpha-\alpha_0\right)\right)^2\geq c_1\left|\alpha-\alpha_0\right|_2^2,$$

due to Assumption 3.1 (iii).

Also, note that

$$\begin{aligned}
&\left|\mathbb{E}\left(x_t'\delta_0\left(1_t\left(\gamma\right)-1_t\left(\gamma_0\right)\right)\right)Z_t\left(\gamma\right)'\left(\alpha-\alpha_0\right)\right|\\
&\leq T^{-\varphi}\mathbb{E}\Big[\left|x_t'd_0\right|\left|1_t\left(\gamma\right)-1_t\left(\gamma_0\right)\right|\left|Z_t\left(\gamma\right)\right|_2\left|\alpha-\alpha_0\right|_2\Big]\\
&\leq 2T^{-\varphi}|d_0|_2C_0C_1\left|\gamma-\gamma_0\right|_2\left|\alpha-\alpha_0\right|_2,
\end{aligned}$$

where the second inequality comes from Assumption 3.1 (i) and Assumption A.1 (i). Combining the inequalities above together yields that

$$\begin{aligned}
R\left(\alpha,\gamma\right)&=\mathbb{E}\left(Z_t\left(\gamma\right)'\left(\alpha-\alpha_0\right)\right)^2+\mathbb{E}\left(x_t'\delta_0\left(1_t\left(\gamma\right)-1_t\left(\gamma_0\right)\right)\right)^2\\
&\quad+2\mathbb{E}\left(x_t'\delta_0\left(1_t\left(\gamma\right)-1_t\left(\gamma_0\right)\right)\right)Z_t\left(\gamma\right)'\left(\alpha-\alpha_0\right)\\
&\geq c_1\left|\alpha-\alpha_0\right|_2^2+c_0T^{-2\varphi}\left|\gamma-\gamma_0\right|_2-C_2T^{-\varphi}\left|\gamma-\gamma_0\right|_2\left|\alpha-\alpha_0\right|_2,
\end{aligned}\tag{F.8}$$

where $C_2=2|d_0|_2C_0C_1$.

We consider two cases: (i) $c_1\left|\alpha-\alpha_0\right|_2\geq 2C_2T^{-\varphi}\left|\gamma-\gamma_0\right|_2$ and (ii) $c_1\left|\alpha-\alpha_0\right|_2<2C_2T^{-\varphi}\left|\gamma-\gamma_0\right|_2.$

When (i) holds,

$$R(\alpha,\gamma) \geq \frac{c_1}{2}|\alpha-\alpha_0|_2^2 + c_0 T^{-2\varphi}|\gamma-\gamma_0|_2 .$$

When (ii) holds, we have that

$$C_2 T^{-\varphi}|\gamma-\gamma_0|_2 |\alpha-\alpha_0|_2 < 2c_1^{-1}C_2^2 T^{-2\varphi}|\gamma-\gamma_0|_2^2 .$$

Then under (ii),

$$\begin{aligned} & c_0 T^{-2\varphi}|\gamma-\gamma_0|_2 - C_2 T^{-\varphi}|\gamma-\gamma_0|_2 |\alpha-\alpha_0|_2 \\ & > T^{-2\varphi}|\gamma-\gamma_0|_2 \left[c_0 - 2c_1^{-1}C_2^2|\gamma-\gamma_0|_2\right]. \end{aligned}$$

Thus, as long as $|\gamma-\gamma_0|_2 \leq c_0c_1/(4C_2^2)$, we obtain the desired result. This completes the proof of Step 1 by taking $c = \min\{c_0, c_1\}/2$ since $|\widehat{\gamma}-\gamma_0|_2 = o_P(1)$ by Lemma F.1. ∎

Proof of Step 2. To prove (F.5), note that as in (F.4),

$$\begin{aligned} & \frac{1}{2}|\mathbb{G}_T(\alpha,\gamma) - \mathbb{G}_T(\alpha_0,\gamma_0)| \qquad \text{(F.9)} \\ & \leq \left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t Z_t(\gamma)'(\alpha-\alpha_0)\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\delta_0(1_t(\gamma)-1_t(\gamma_0))\right| \\ & = O_P\left(\frac{1}{\sqrt{T}}\right)|\alpha-\alpha_2|_2 + \eta T^{-2\varphi}|\gamma-\gamma_0|_2 + O_P\left(\frac{1}{T}\right) \end{aligned}$$

for any $0 < \eta < c/2$, by the MDS CLT and Lemma I.1 for the first term $T^{-1/2}\sum_{t=1}^{T}\varepsilon_t Z_t(\gamma)$ and by Assumption F.2 for the second term $T^{-1}\sum_{t=1}^{T}\varepsilon_t x_t'\delta_0(1_t(\gamma)-1_t(\gamma_0))$.

We now prove (F.6). Note that for any $0 < \eta < c/2$, as in (F.3),

$$\begin{aligned} & |\mathbb{R}_T(\alpha,\gamma) - R(\alpha,\gamma)| \qquad \text{(F.10)} \\ & \leq \left|(\alpha-\alpha_0)'\frac{1}{T}\sum_{t=1}^{T}\left(Z_t(\gamma)Z_t(\gamma)' - \mathbb{E}Z_t(\gamma)Z_t(\gamma)'\right)(\alpha-\alpha_0)\right| \\ & + \left|\frac{1}{T}\sum_{t=1}^{T}\left(x_t'\delta_0\right)^2|1_t(\gamma)-1_t(\gamma_0)| - \mathbb{E}\left(x_t'\delta_0\right)^2|1_t(\gamma)-1_t(\gamma_0)|\right| \\ & + \left|\frac{2}{T}\sum_{t=1}^{T}\delta_0'\left[x_t(1_t(\gamma)-1_t(\gamma_0))Z_t(\gamma) - \mathbb{E}\left[x_t(1_t(\gamma)-1_t(\gamma_0))Z_t(\gamma)\right]\right]'(\alpha-\alpha_0)\right| \\ & \leq o_P\left(|\alpha-\alpha_0|_2^2\right) + O_P\left(\frac{1}{T}\right) + \eta T^{-2\varphi}|\gamma-\gamma_0|_2 \end{aligned}$$

by ULLN for the first term and by Lemma F.2 for the second and third terms. This completes the proof. ∎

F.3 Asymptotic Distribution

Proof of Theorem 3.1. Let $r_T \equiv T^{1-2\varphi}$, $a \equiv \sqrt{T}(\alpha - \alpha_0)$ and $g \equiv r_T(\gamma - \gamma_0)$. To prove the theorem, we first derive the weak convergence of the process

$$\mathbb{K}_T(a,g) \equiv T\left(\mathbb{S}_T\left(\alpha_0 + a \cdot T^{-1/2}, \gamma_0 + g \cdot r_T^{-1}\right) - \mathbb{S}_T(\alpha_0, \gamma_0)\right),$$

over an arbitrary compact set, say $\mathcal{AG}$, and then apply the argmax continuous mapping theorem to obtain the limit distribution of $\widehat{\alpha}$ and $\widehat{\gamma}$.

Step 1. The following decomposition holds uniformly in $(a,g) \in \mathcal{AG}$:

$$\mathbb{K}_T(a,g) = \mathbb{K}_{1T}(a) + \mathbb{K}_{2T}(g) - 2\mathbb{K}_{3T}(g) + o_P(1),$$

where

$$\begin{aligned}
\mathbb{K}_{1T}(a) &:= a'\mathbb{E}Z_t(\gamma_0)Z_t(\gamma_0)'a - \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t Z_t(\gamma_0)'a, \\
\mathbb{K}_{2T}(g) &:= T \cdot \mathbb{E}\left[\left(x_t'\delta_0\right)^2\left|1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t\right|\right], \\
\mathbb{K}_{3T}(g) &:= \sum_{t=1}^{T}\varepsilon_t x_t'\delta_0\left(1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t\right).
\end{aligned}$$

Proof of Step 1. To begin with, note that (F.10) and Lemma F.2 together imply that

$$\begin{aligned}
&T \cdot \left[\mathbb{R}_T\left(\alpha_0 + a \cdot T^{-1/2}, \gamma_0 + g \cdot r_T^{-1}\right) - R\left(\alpha_0 + a \cdot T^{-1/2}, \gamma_0 + g \cdot r_T^{-1}\right)\right] \\
&= o_P(1) \text{ uniformly in } (a,g) \in \mathcal{AG}.
\end{aligned} \tag{F.11}$$

Recall (F.8) and write that

$$\begin{aligned}
&T \cdot R\left(\alpha_0 + a \cdot T^{-1/2}, \gamma_0 + g \cdot r_T^{-1}\right) \\
&= a'\mathbb{E}\left[Z_t\left(\gamma_0 + g \cdot r_T^{-1}\right)Z_t\left(\gamma_0 + g \cdot r_T^{-1}\right)'\right]a \\
&+ T \cdot \mathbb{E}\left(x_t'\delta_0\right)^2\left|1\left\{f_t'\left(\gamma_0 + g \cdot r_T^{-1}\right) > 0\right\} - 1\left\{f_t'\gamma_0\right\}\right| \\
&+ 2T^{1/2} \cdot \mathbb{E}\left(x_t'\delta_0\left(1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t(\gamma_0)\right)\right)Z_t\left(\gamma_0 + g \cdot r_T^{-1}\right)'a.
\end{aligned} \tag{F.12}$$

Then, due to Assumption 3.2,

$$
\begin{aligned}
a'\Big\{\mathbb{E}\left[Z_t\left(\gamma_0+g\cdot r_T^{-1}\right)Z_t\left(\gamma_0+g\cdot r_T^{-1}\right)'\right]-\mathbb{E}\left[Z_t\left(\gamma_0\right)Z_t\left(\gamma_0\right))'\right]\Big\}\,a=o_P(1),\\
T^{1/2}\cdot\mathbb{E}\left[\left(x_t'\delta_0\left(1_t\left(\gamma_0+g\cdot r_T^{-1}\right)-1_t\left(\gamma_0\right)\right)\right)Z_t\left(\gamma_0+g\cdot r_T^{-1}\right)'\right]a=o_P(1)
\end{aligned}
\tag{F.13}
$$

uniformly in $(a,g)\in\mathcal{AG}$. Then combining (F.11)-(F.13) yields that

$$
\begin{aligned}
&T\cdot\mathbb{R}_T\left(\alpha_0+a\cdot T^{-1/2},\gamma_0+g\cdot r_T^{-1}\right)\\
&=a'\mathbb{E}\left[Z_t\left(\gamma_0\right)Z_t\left(\gamma_0\right)'\right]a+T\cdot\mathbb{E}\left(x_t'\delta_0\right)^2\left|1\left\{f_t'\left(\gamma_0+g\cdot r_T^{-1}\right)>0\right\}-1\left\{f_t'\gamma_0\right\}\right|\\
&+o_P(1)\ \text{ uniformly in } (a,g)\in\mathcal{AG}.
\end{aligned}
\tag{F.14}
$$

We now consider the term $T\left[\mathbb{G}_T\left(\alpha_0+a\cdot T^{-1/2},\gamma_0+g\cdot r_T^{-1}\right)-\mathbb{G}_T\left(\alpha_0,\gamma_0\right)\right]$. First, note that due to Lemma I.1,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\left[Z_t\left(\gamma_0+g\cdot r_T^{-1}\right)-Z_t\left(\gamma_0\right)\right]'a=o_P\left(1\right)
\tag{F.15}
$$

uniformly in $(a,g)\in\mathcal{AG}$. Then, recall (F.4) and write that

$$
\begin{aligned}
&T\left[\mathbb{G}_T\left(\alpha_0+a\cdot T^{-1/2},\gamma_0+g\cdot r_T^{-1}\right)-\mathbb{G}_T\left(\alpha_0,\gamma_0\right)\right]\\
&=\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_tZ_t\left(\gamma_0+g\cdot r_T^{-1}\right)'a+2\sum_{t=1}^{T}\varepsilon_tx_t'\delta_0\left(1_t\left(\gamma_0+g\cdot r_T^{-1}\right)-1_t\left(\gamma_0\right)\right)\\
&=\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_tZ_t\left(\gamma_0\right)'a+2\sum_{t=1}^{T}\varepsilon_tx_t'\delta_0\left(1_t\left(\gamma_0+g\cdot r_T^{-1}\right)-1_t\left(\gamma_0\right)\right)+o_P\left(1\right),
\end{aligned}
\tag{F.16}
$$

uniformly in $(a,g)\in\mathcal{AG}$, where the last equality follows from (F.15). Then Step 1 follows immediately recalling the decomposition in (F.2) and collecting the leading terms in (F.14) and (F.16). ∎

In view of Step 1, the limiting distribution of a is determined by $\mathbb{K}_{1T}\left(a\right)$. That is,

$$
a=\left[\mathbb{E}Z_t\left(\gamma_0\right)Z_t\left(\gamma_0\right)'\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_tZ_t\left(\gamma_0\right)+o_P(1).
$$

Then the first desired result follows directly from the martingale difference central limit theorem (e.g. Hall and Heyde, 1980).

Step 2.

$$
T^{1-2\varphi}\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\underset{g\in\mathcal{G}}{\operatorname{argmin}}\,\mathbb{E}\left[\left(x_t'd_0\right)^2\left|f_t'g\right|p_{u_t|f_{2t}}(0)\right]+2W\left(g\right),
$$

where W is a Gaussian process whose covariance kernel is given by $H(\cdot,\cdot)$ in (3.1) and $\mathcal{G} = \{g \in \mathbb{R}^d : g_1 = 0\}$.

Proof of Step 2. The distribution of g is determined by $\mathbb{K}_{2T}(g) - 2\mathbb{K}_{3T}(g)$. For the weak convergence of $\mathbb{K}_{3T}(g)$, we need to verify the tightness of the process and the finite dimensional convergence. The tightness is the consequence of Lemma I.1 since for any finite g and for any $c > 0$,

$$\begin{aligned} & \mathbb{P}\left\{ \sup_{|h-g|<\epsilon} |\mathbb{K}_{3T}(g) - \mathbb{K}_{3T}(h)| > c \right\} \\ = \; & \mathbb{P}\left\{ \sup_{|\bar{\gamma}-\gamma|<\epsilon/r_T} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t x_t' d_0 \left(1_t(\bar{\gamma}) - 1_t(\gamma)\right) \right| > \frac{c}{2\sqrt{T}} T^{\varphi} \right\} \\ \leq \; & C\frac{\epsilon^2}{c^4}, \end{aligned}$$

which can be made arbitrarily small by choosing ϵ small. For the fidi, we apply the martingale difference central limit theorem (e.g. Hall and Heyde, 1980). Specifically, let $w_t = \sqrt{r_T}\varepsilon_t x_t' d_0 \left(1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t\right)$ and verify that $\max_t |w_t| = o_P\left(\sqrt{T}\right)$ and that $\frac{1}{T}\sum_{t=1}^{T} w_t^2$ has a proper non-degenerate probability limit. However, $T^{-2}\mathbb{E}\max_t w_t^4 \leq T^{-1}\mathbb{E}w_t^4$ since $\max_t |a_t| \leq \sum_{t=1}^{T} |a_t|$ and w_t is stationary. Now,

$$T^{-1}\mathbb{E}w_t^4 = T^{-1}r_T^2\mathbb{E}\left[\left(\varepsilon_t x_t' d_0\right)^4 \left|1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t\right|\right] \leq CT^{-1}r_T = o(1).$$

Furthermore, $\frac{1}{T}\sum_{t=1}^{T}\left(w_t^2 - \mathbb{E}w_t^2\right) = o_P(1)$. The limit of $\mathbb{E}w_t^2$ will be given later while we characterize the covariance kernel of the process $\mathbb{K}_{3T}(g)$.

To derive the covariance kernel of $\mathbb{K}_{3T}(g)$ and the limit of $\mathbb{K}_{2T}(g)$, we need to derive the limit of the type

$$\lim_{m\to\infty} m\mathbb{E}\eta_t^2 \left|1\left\{f_t'(\gamma_0 + s/m) > 0\right\} - 1\left\{f_t'(\gamma_0 + g/m) > 0\right\}\right|$$

for some random variable η_t given $s \neq g$. We split the remainder of the proof into two cases.

Remark F.1. In the meantime, we note that this proof also implies that the covariance between the second term in $\mathbb{K}_{1T}(a)$ and $\mathbb{K}_{3T}(g)$ degenerates, which implies the asymptotic independence between two processes.

Recall that $\gamma_1 = 1$. With this normalization, we need to fix the first element of g in $\mathbb{K}_{2T}(g)$ and $\mathbb{K}_{3T}(g)$ at zero. Thus, we assume $g \in \mathbb{R}^{d-1}$ with a slight abuse of notation and introduce $u_t = f_t'\gamma_0$ and

$$h\left((\eta_t, u_t, f_{2t}), g/m\right) = \eta_t 1\left\{u_t + f_{2t}' g/m > 0\right\}$$

for $g \in \mathbb{R}^{d-1}$ and some random variable η_t, which will be made more explicit later. Then, the asymptotic covariances of the process $\mathbb{K}_{3T}(g)$ and the limit of $\mathbb{K}_{2T}(g)$ are characterized by the limit of the type

$$L(s,g) = \lim_{m\to\infty} m\mathbb{E}\left(h(\cdot, s/m) - h(\cdot, g/m)\right)^2,$$

for $g, s \in \mathbb{R}^{d-1}$. That is, for the asymptotic covariance kernel $H(s,g)$ of $\mathbb{K}_{3T}(g)$, set $\eta_t = x_t' d_0 \varepsilon_t$, which is a martingale difference sequence to render $\mathbb{E}h(\cdot, g/m) = 0$, and $m = T^{1-2\varphi}$. Then,

$$\begin{aligned}
H(s,g) &= \operatorname{cov}\left(\mathbb{K}_{3T}(s), \mathbb{K}_{3T}(g)\right) \\
&= \mathbb{E}\left(\left(h(\cdot, s/m) - \eta_t 1\{u_t > 0\}\right)\left(h(\cdot, g/m) - \eta_t 1\{u_t > 0\}\right)\right) \\
&= \frac{1}{2}\left(L(s,0) + L(g,0) - L(s,g)\right),
\end{aligned}$$

since $2ab = a^2 + b^2 - (a-b)^2$ and $h(\cdot, 0) = \eta_t 1\{u_t > 0\}$. On the other hand, the limit of $\mathbb{K}_{2T}(g)$ will be given by $L(g,0)$ with $\eta_t = x_t' d_0$.

Note that

$$\begin{aligned}
L(s,g) &= \lim_{m\to\infty} m\mathbb{E}\eta_t^2 \left|1\left\{u_t + f_{2t}' s/m > 0\right\} - 1\left\{u_t + f_{2t}' g/m > 0\right\}\right| \\
&= m\mathbb{E}\eta_t^2 1\left\{u_t + f_{2t}' s/m > 0 \geq u_t + f_{2t}' g/m\right\} \\
&\quad + m\mathbb{E}\eta_t^2 1\left\{u_t + f_{2t}' g/m > 0 \geq u_t + f_{2t}' s/m\right\}.
\end{aligned}$$

Furthermore, let $p_{u|f_2}(\cdot)$ and P_2 denote the conditional density of u_t given $f_{2t} = f_2$ and the probability measure for f_{2t}, respectively, and note that

$$\begin{aligned}
& m\mathbb{E}\eta_t^2 1\left\{u_t + f_{2t}' s/m > 0 \geq u_t + f_{2t}' g/m\right\} \\
= & \int\int \mathbb{E}\left[\eta_t^2 | w/m, f_2\right] 1\left\{-f_2' g \geq w > -f_2' s\right\} p_{u|f_2}(w/m)\, dw dP_2 \\
\to & \int \mathbb{E}\left[\eta_t^2 | 0, f_2\right]\left(-f_2' g + f_2' s\right) 1\left(f_2' g < f_2' s\right) p_{u|f_2}(0)\, dP_2,
\end{aligned}$$

where the equality is by a change of variables, $w = m \cdot u$ and the convergence is as $m \to \infty$ by the dominated convergence theorem (DCT). This implies that

$$L(s,g) = \int \mathbb{E}\left[\eta_t^2 | 0, f_2\right]\left|f_2' g - f_2' s\right| p_{u|f_2}(0)\, dP_2.$$

In the special case where $z_t' g < 0 < z_t' s$ almost surely, $L(s,g) = L(s,0) + L(g,0)$. This happens when $f_t = (q_t, -1)$ and thus z_t is a constant given u_t.

Therefore, putting together,

$$T^{1-2\varphi}\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\underset{g\in\mathbb{R}^d:g_1=0}{\operatorname{argmin}}\ \mathbb{E}\left[\left(x_t'd_0\right)^2\left|f_t'g\right|p_{u_t|f_{2t}}(0)\right]+2W\left(g\right),$$

where W is a Gaussian process whose covariance kernel is given by

$$H\left(s,g\right)=\frac{1}{2}\mathbb{E}\left[\left(x_t'd_0\right)^2\left(\left|f_t'g\right|+\left|f_t's\right|-\left|f_t'\left(g-s\right)\right|\right)p_{u_t|f_{2t}}(0)\right].$$

∎

Step 3. Asymptotically, a and g are independent of each other.

Proof of Step 3. This is straightforward due to the separability of $\mathbb{K}$ into functions of a and g, and due to Remark F.1 that addresses the independence between the processes of a and g. ∎

G Proof of Asymptotics in Section 4: Estimated Factors

G.1 A Roadmap of the Proof

Due to the complexity of the proof, we begin with a roadmap to help readers follow the steps of the proof.

Step I. We first prove a probability bound for $|\widetilde{f}_t-\widehat{f}_t|_2$ in Section G.3.1, where

$$\widehat{f}_t=H_T'g_t+H_T'\frac{h_t}{\sqrt{N}}.$$

Step II. We then replace the PCA estimator $\widetilde{f}_t$ in the objective function $\widetilde{\mathbb{S}}_T\left(\alpha,\gamma\right)$ with its first-order approximation $\widehat{f}_t$, and show that the effect of such a replacement is negligible for the convergence rates of the estimators we obtain in the later steps in Section G.3.3.

Step III. We show the consistency of estimators. To do so and to derive the convergence rates in the later steps, we use the alternative parametrization $\phi=H_T\gamma$, which helps us derive various uniform convergence lemmas. Note that the reparametrization is fine for the consistency and convergence rate results of the original parameter estimates since H_T is nonsingular with probability approaching one.

Step IV. We then decompose the objective function into the following form:

$$\widetilde{\mathbb{S}}_T\left(\alpha,H_T^{-1}\phi\right)-\widetilde{\mathbb{S}}_T\left(\alpha_0,H_T^{-1}\phi_0\right)=\mathbb{R}_T(\alpha,\phi)+\mathbb{G}_1(\phi)-\mathbb{C}(\alpha,\phi),$$

where $\mathbb{R}_T(\cdot,\cdot)$ and $\mathbb{G}_1(\cdot)$ are deterministic functions and $\mathbb{C}(\cdot,\cdot)$ is a stochastic function. The formal definitions are given before Lemma G.3. Then as $\widetilde{\mathbb{S}}_T(\widehat{\alpha},\widehat{\gamma})-\widetilde{\mathbb{S}}_T(\alpha_0,\gamma_0)\leq 0$, the decomposition yields: for $\widehat{\phi}=H_T\widehat{\gamma}$,

$$C|\widehat{\alpha}-\alpha_0|_2^2+\mathbb{G}_1(\widehat{\phi})\leq\mathbb{C}(\widehat{\alpha},\widehat{\phi}) \tag{G.1}$$

where $\mathbb{R}_T(\alpha,\phi)$ is lower bounded by $C|\alpha-\alpha_0|_2^2$ uniformly. Then, Lemmas G.3 and G.4 establish uniform stochastic upper bounds for $\mathbb{C}(\widehat{\alpha},\widehat{\phi})$ through maximal inequalities.

Step V. Next, we derive a uniform lower bound for $\mathbb{G}_1(\phi)$ over ϕ near ϕ_0 and over the ratio $\sqrt{N}/T^{1-2\varphi}$ in Lemma G.5. In particular, $\mathbb{G}_1(\phi)$ has a "kink" lower bound:

$$\mathbb{G}_1(\phi)\geq CT^{-2\varphi}|\phi-\phi_0|_2-\frac{C}{\sqrt{N}T^{2\varphi}}.$$

These bounds lead to the rate of convergence:

$$|\widehat{\alpha}-\alpha_0|_2=O_P(T^{-1/2}+N^{-1/4}T^{-\varphi}),\quad |\widehat{\phi}-\phi_0|_2=O_P(T^{-(1-2\varphi)}+N^{-1/2}).$$

These bounds and the rates are sharp in the case $\sqrt{N}/T^{1-2\varphi}\to\infty$, and are identical to the case of the known factor.

Step VI. It turns out the lower and upper bounds for $\mathbb{G}_1(\cdot)$ and $\mathbb{C}(\cdot)$ are not sharp when $\sqrt{N}/T^{1-2\varphi}\to\omega<\infty$. We then provide sharper bounds for these terms. In particular, obtaining the sharp lower bound for $\mathbb{G}_1(\cdot)$ is most challenging and involves complicated expansions. We establish in Lemma G.6 that it has a quadratic lower bound with an unusual error rate:

$$\mathbb{G}_1(\phi)\quad\geq\quad CT^{-2\varphi}\sqrt{N}|\phi-\phi_0|_2^2-O(\frac{1}{T^{2\varphi}N^{5/6}}).$$

These lead to a sharp rate for $\widehat{\phi},\widehat{\gamma}$ in Proposition G.4 in the case of $\omega<\infty$.

Step VII. Finally, we derive the limiting distributions for $\widehat{\alpha}$ and $\widehat{\gamma}$. This involves utilizing the convergence rates we obtained through the preceding steps to recenter, rescale and reparametrize the original criterion function, which is parametrized not by ϕ but by γ. Then, we establish the stochastic equicontinuity of the empirical process part of the transformed process (i.e. centered process) in Section G.7.1 and the careful expansion of the drift (i.e. bias) part of the process as a function of the limit $\omega=\lim_{N,T}\sqrt{N}T^{-1+2\varphi}$ in Section G.7.2. Due to the random rotation matrix H_T incurred by the factor estimation, we prove an extended continuous mapping theorem in Lemma I.4, to derive the weak convergence of the transformed criterion function. The remaining step is the application of the argmax continuous mapping theorem. The new CMT extends Theorem 1.11.1 of van der Vaart and Wellner (1996) to allowing stochastic drifting functions $\mathbb{G}_n$ (while van der Vaart and Wellner (1996) requires $\mathbb{G}_n$ be deterministic).

G.2 Discussion on Assumption 4.5

We discuss the reasons why Assumption 4.5 presents various conditions on several different conditional distributions and why those conditional distributions are well defined. A key technical issue in expanding the least squares loss function, in the unknown factor case, is to consider the properties of the conditional density of $g_t'\phi_0$, given $g_t'(\phi-\phi_0)$ and (x_t, h_t). It is needed in bounding terms of the form:

$$\mathbb{E}\left[(x_t'\delta_0)^2\Psi(h_t'\phi_0, g_t'\phi_0, g_t'(\phi-\phi_0)\right]$$

with a suitably defined function Ψ. But we should be cautious that such a conditional density might be degenerated because given $g_t'(\phi-\phi_0)$, there might be no degree of freedom left for $g_t'\phi_0$. To address this issue, we observe that by the identification condition, we can write $\gamma = (1, \gamma_2) = H_T^{-1}\phi$, where 1 is the first element of γ. Let the corresponding factor be $f_t = (f_{1t}, f_{2t})$. Then $g_t'(\phi-\phi_0) = f_t'(\gamma-\gamma_0) = f_{2t}'(\gamma_2-\gamma_{02})$, so it depends on f_t only through f_{2t}. As such, we can consider the conditional density of $f_t'\gamma_0$ given (f_{2t}, x_t, h_t). Being given f_{2t} still leaves degrees of freedom for $f_t'\gamma_0$, so such conditional density is well defined.

In the lower bound for $\mathbb{G}_1(\phi)$ in Step VI, the problem eventually reduces to lower bounding

$$\mathbb{E}\left[(x_t'\delta_0)^2 p_{f_t'\gamma_0|f_{2t},x_t,h_t}(0)|g_t'(\phi-\phi_0)|^2 1\{|g_t|_2 < M_0\}\right]$$

for a sufficiently large M_0. We can apply the above argument to achieve a tight quadratic lower bound $C|\phi-\phi_0|_2^2$, so long as the conditional density $p_{f_t'\gamma_0|f_{2t},x_t,h_t}(0)$ and the eigenvalues of $\mathbb{E}[(x_t'd_0)^2|g_t, h_t]$ are bounded away from zero. In addition, here we also need to upper bound $\mathbb{P}(\frac{h_t'\phi}{\sqrt{N}} < g_t'(\phi-\phi_0) < \frac{h_t'\phi_0}{\sqrt{N}}|h_t)$ and $\mathbb{P}(\frac{h_t'\phi}{\sqrt{N}} < g_t'\phi < \frac{h_t'\phi_0}{\sqrt{N}}|h_t)$. This is ensured by the condition $\sup_{|u|<c} p_{g_t'r|x_t,h_t}(u) \leq M$.

When we derive a lower bound for $\mathbb{G}_1(\phi)$ in Step V, we also need such an argument for the conditional density of $\widehat{f}_t = H_T'\breve{g}_t$, where $\breve{g}_t = g_t + \frac{h_t}{\sqrt{N}}$ is the perturbed factors, estimated by the PCA. For instance, we need a lower bound when $\Psi = \mathbb{P}\left(0 < \breve{g}_t'\phi_0 < |\breve{g}_t'(\phi-\phi_0)|\right)$. To derive this lower bound, write $\widehat{f}_t = (\widehat{f}_{1t}, \widehat{f}_{2t})$. Then $\breve{g}_t'(\phi-\phi_0)$ depends on $\widehat{f}_t$ only through $\widehat{f}_{2t}$. As such, we can consider the conditional density of $\widehat{f}_t'\gamma_0$ given $(\widehat{f}_{2t}, x_t)$, and obtain a lower bound

$$\mathbb{E}\left[(x_t'd_0)^2 1(0 < \breve{g}_t'\phi_0 < |\breve{g}_t'(\phi-\phi_0)|)\right] \geq \inf_{m,x,\widehat{f}_{2t}} p_{\widehat{f}_t'\gamma_0|\widehat{f}_{2t},x_t}(m)\mathbb{E}\left[|\breve{g}_t'(\phi-\phi_0)|\right] \geq C|\phi-\phi_0|_2,$$

where it is assumed that $\inf_{|m|<c}\inf_{x,\widehat{f}_{2t}} p_{\widehat{f}_t'\gamma_0|\widehat{f}_{2t},x_t}(m)\geq c_0>0$. The need for arguments like this gives rise to Assumption 4.5 (i)-(iv).

G.3 Consistency

G.3.1 A probability bound for $|\widetilde{f}_t-\widehat{f}_t|_2$

The stochastic order of the approximation error of $\widetilde{f}_t-\widehat{f}_t$ has been well studied in the literature (see, e.g. Bai, 2003). However, all the existing results in the literature are on the rates of convergence for $\widetilde{f}_t-\widehat{f}_t$ of a fixed t and for $\frac{1}{T}\sum_t|\widetilde{f}_t-\widehat{f}_t|_2^2$. We strengthen these results below by obtaining the following probability bound.

Proposition G.1. *Suppose $T=O(N)$. Define*

$$\Delta_f=\frac{(\log T)^{2/c_1}}{T}$$

Then for a sufficiently large constant $C>0$, and $\widehat{f}_t=H_T'(g_t+\frac{h_t}{\sqrt{N}})$,

$$\mathbb{P}(|\widetilde{f}_t-\widehat{f}_t|_2>C\Delta_f)\leq O(T^{-6}).$$

Proof of Proposition G.1. The proof consists of several steps. Recall that $\widetilde{f}_{1t}$ denotes the $K\times 1$ vector of PCA estimator of g_{1t}. Write $e_t=(e_{1t},...,e_{Nt})'$.

Step 1: Decomposition of $\widetilde{f}_t-H_T'g_t$

Define $K\times K$ matrix $\widetilde{H}_T'=V_T^{-1}\frac{1}{T}\sum_{t=1}^T\widetilde{f}_{1t}g_{1t}'S_\Lambda$, and $S_\Lambda=\frac{1}{N}\Lambda'\Lambda$. Also let V_T be the $K\times K$ diagonal matrix whose entries are the first K eigenvalues of $\mathcal{Y}\mathcal{Y}'/NT$ (equivalently, the first K eigenvalues of $\frac{1}{NT}\sum_t\mathcal{Y}_t\mathcal{Y}_t'$). We have

$$\widetilde{f}_{1t}-\widetilde{H}_T'g_{1t}=\widetilde{H}_T'S_\Lambda^{-1}\frac{1}{N}\Lambda'e_t+\sum_{d=1}^6A_{t,d}, \tag{G.2}$$

where

$$\begin{aligned}
A_{t,1} &= V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^T g_{1s}\frac{1}{N}\mathbb{E}e_s'e_t,\\
A_{t,2} &= V_T^{-1}\frac{1}{T}\sum_{s=1}^T(\widetilde{f}_{1s}-\widetilde{H}_T'g_{1s})\frac{1}{N}\mathbb{E}e_s'e_t,\\
A_{t,3} &= V_T^{-1}\frac{1}{T}\sum_{s=1}^T(\widetilde{f}_{1s}-\widetilde{H}_T'g_{1s})\frac{1}{N}(e_s'e_t-\mathbb{E}e_s'e_t),\\
A_{t,4} &= V_T^{-1}\widetilde{H}_T'\frac{1}{TN}\sum_{s=1}^T g_{1s}(e_s'e_t-\mathbb{E}e_s'e_t),
\end{aligned}$$

$$
\begin{aligned}
A_{t,5} &= V_T^{-1}\frac{1}{T}\sum_{s=1}^{T}(\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s}) g_{1t}' \frac{1}{N}\sum_{i=1}^{N}\lambda_i e_{is}, \\
A_{t,6} &= V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^{T} g_{1s} g_{1t}' \frac{1}{N}\sum_{i=1}^{N}\lambda_i e_{is}.
\end{aligned}
$$

Hence for $H_T' = \text{diag}\{\widetilde{H}_T', 1\}$, $g_t = (g_{1t}', 1)'$, $\widetilde{f}_t = (\widetilde{f}_{1t}', 1)'$, $h_t = (S_\Lambda^{-1}\frac{\Lambda' e_t}{\sqrt{N}}, 0)'$, and $\widehat{f}_t = H_T'(g_t + \frac{h_t}{\sqrt{N}})$, we have

$$
\widetilde{f}_t - \widehat{f}_t = (\sum_{d=1}^{6} A_{t,d}, 0)'. \tag{G.3}
$$

Step 2: Bounding $\frac{1}{T}\sum_t |\widetilde{f}_{1t} - \widetilde{H}_T' g_{1t}|_2^2$

Note that

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}|\widetilde{f}_{1t} - \widetilde{H}_T' g_{1t}|_2^2 \leq\ & 4\frac{1}{T}\sum_{t=1}^{T}|\widetilde{H}_T'\frac{h_t}{\sqrt{N}}|_2^2 + 4\frac{1}{T}\sum_{t=1}^{T}|V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^{T} g_{1s}\frac{1}{N}\mathbb{E}e_s' e_t|_2^2 \\
& + \frac{1}{T}\sum_{s=1}^{T}|\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s}|_2^2 (a_1 + a_2 + a_3) \\
& + 8\frac{1}{T}\sum_{t=1}^{T}|V_T^{-1}\widetilde{H}_T'\frac{1}{TN}\sum_{s=1}^{T} g_{1s}(e_s' e_t - \mathbb{E}e_s' e_t)|_2^2 \\
& + 8\frac{1}{T}\sum_{t=1}^{T}|V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\sum_{i=1}^{N} g_{1s} e_{is}\lambda_i' g_{1t}|_2^2,
\end{aligned}
$$

where

$$
a_1 = |V_T^{-1}|_2^2\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}|\frac{1}{N}(e_s' e_t - \mathbb{E}e_s' e_t)|_2^2, \quad a_2 = |V_T^{-1}|_2^2\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}|g_{1t}'\frac{1}{N}\Lambda' e_s|_2^2
$$

and assuming $\frac{1}{NT}\sum_{t,s\leq T}\sum_{i\leq N}|\mathbb{E}e_{it}e_{is}| < C$,

$$
a_3 = |V_T^{-1}|_2^2\max_{s,t}|\frac{1}{N}\mathbb{E}e_s' e_t|\frac{1}{T^2}\sum_t\sum_{s=1}^{T}|\frac{1}{N}\mathbb{E}e_s' e_t| \leq C|V_T^{-1}|_2^2\frac{1}{T}.
$$

Hence for $c_{NT} = (1 - a_1 - a_2 - a_3)$,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}|\widetilde{f}_{1t} - \widetilde{H}_T' g_{1t}|_2^2 c_{NT} \leq\ & 8\frac{1}{T}\sum_{t=1}^{T}|V_T^{-1}\widetilde{H}_T'\frac{1}{TN}\sum_{s=1}^{T} g_{1s}(e_s' e_t - \mathbb{E}e_s' e_t)|_2^2 \\
& + 4\frac{1}{T}\sum_{t=1}^{T}|\widetilde{H}_T'\frac{h_t}{\sqrt{N}}|_2^2 + 4\frac{1}{T}\sum_{t=1}^{T}|V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^{T} g_{1s}\frac{1}{N}\mathbb{E}e_s' e_t|_2^2
\end{aligned}
$$

$$+8\frac{1}{T}\sum_{t=1}^{T}|V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\sum_{i=1}^{N}g_{1s}e_{is}\lambda_i'g_{1t}|_2^2. \quad \text{(G.4)}$$

Next we provide probability bounds for each term on the right hand side below.

Step 3: Proving that $T^6\mathbb{P}(|V_T^{-1}|_2 > C_v) + T^6\mathbb{P}(|\widetilde{H}_T|_2 > C_H) = o(1)$ for some $C_v, C_H > 0$

Let V be the diagonal matrix consisting of the first K eigenvalues of $\Sigma_\Lambda^{1/2}\mathbb{E}[g_{1t}g_{1t}']\Sigma_\Lambda^{1/2}$. On the event $|V_T - V|_2 < \lambda_{\min}(V)/2$,

$$|V_T^{-1}|_2 = \lambda_{\min}^{-1}(V_T) \le 2\lambda_{\min}^{-1}(V) \le 2\lambda_{\min}^{-1}(\frac{1}{N}\Lambda'\Lambda)\lambda_{\min}^{-1}(\mathbb{E}g_{1t}g_{1t}') < C_v.$$

We now show $T^6\mathbb{P}(|V_T - V|_2 > \lambda_{\min}(V)/2) = o(1)$. By Weyl's theorem,

$$\begin{aligned}
|V_T - V|_2 &\le |\frac{1}{NT}\sum_t \mathcal{Y}_t\mathcal{Y}_t' - \frac{1}{N}\Lambda\mathbb{E}g_{1t}g_{1t}'\Lambda'|_2 \le |\frac{1}{N}\Lambda(\mathbb{E}g_{1t}g_{1t}' - \frac{1}{T}\sum_t g_{1t}g_{1t}')\Lambda'|_2 \\
&\quad +2|\frac{1}{N}\Lambda\frac{1}{T}\sum_t g_{1t}e_t'|_2 + |\frac{1}{N}(\frac{1}{T}\sum_t e_te_t' - \mathbb{E}e_te_t')|_2 + \frac{1}{N}|\mathbb{E}e_te_t'|_2 \\
&\le C|\mathbb{E}g_{1t}g_{1t}' - \frac{1}{T}\sum_t g_{1t}g_{1t}'|_2 + C\frac{1}{\sqrt{N}}|\frac{1}{T}\sum_t g_{1t}e_t'|_2 + |\frac{1}{N}(\frac{1}{T}\sum_t e_te_t' - \mathbb{E}e_te_t')|_2 + \frac{C}{N} \\
&= b_1 + b_2 + b_3 + \frac{C}{N}.
\end{aligned}$$

By the Bernstein inequality, for some $M, c, \zeta, r > 0$,

$$\begin{aligned}
T^6\mathbb{P}(b_1 > \lambda_{\min}(V)/9) &= T^6\mathbb{P}(C|Eg_{1t}g_{1t}' - \frac{1}{T}\sum_t g_{1t}g_{1t}'|_2 > \lambda_{\min}(V)/9) \\
&\le T^6\exp(-MT^c) = o(1), \\
T^6\mathbb{P}(b_2 > \lambda_{\min}(V)/9) &= T^6\mathbb{P}(C|\frac{1}{T}\sum_t g_{1t}e_t'|_2 > \sqrt{N}\lambda_{\min}(V)/9) \\
&\le CT^{-3}\max_{i\le N}\mathbb{E}|\frac{1}{\sqrt{T}}\sum_t g_{1t}e_{it}|_2^r \\
&= CT^{-3}\max_i\int_0^\infty \mathbb{P}(|\frac{1}{\sqrt{T}}\sum_t g_{1t}e_{it}|_2 > x^{-r})dx \\
&\le CT^{-3}\int_0^\infty \exp(-Cx^{-\zeta})dx = O(T^{-3}), \\
T^6\mathbb{P}(b_3 > \lambda_{\min}(V)/9) &= T^6\mathbb{P}(|\frac{1}{T}\sum_t e_te_t' - \mathbb{E}e_te_t'|_2 > N\lambda_{\min}(V)/9) \\
&\le CT^{-3}\max_{ij}\mathbb{E}|\frac{1}{\sqrt{T}}\sum_t(e_{it}e_{jt} - \mathbb{E}e_{it}e_{jt})|^r \\
&\le CT^{-3}\max_{ij}\int_0^\infty \mathbb{P}(|\frac{1}{\sqrt{T}}\sum_t(e_{it}e_{jt} - \mathbb{E}e_{it}e_{jt})| > x^{-r})dx \\
&\le CT^{-3}\int_0^\infty \exp(-Cx^{-\zeta})dx = O(T^{-3}).
\end{aligned}$$

Hence

$$
\begin{aligned}
T^6\mathbb{P}(|V_T^{-1}|>C_v) &\leq T^6\mathbb{P}(|V_T^{-1}|_2>C_v, |V_T-V|_2<\lambda_{\min}(V)/2) \\
&\quad +T^6\mathbb{P}(|V_T-V|_2>\lambda_{\min}(V)/2) \\
&= T^6\mathbb{P}(|V_T-V|_2>\lambda_{\min}(V)/2) \\
&\leq T^6\mathbb{P}(b_1+b_2+b_3>\lambda_{\min}(V)/3) \\
&\leq T^6\sum_{i=1}^{3}\mathbb{P}(b_i>\lambda_{\min}(V)/9)=o(1).
\end{aligned}
$$

Now On the event $|V_T^{-1}|_2\leq C_v$, for $C_H>C_\lambda^2C_v(2M_f)^{1/2}K$ (recall $|S_\Lambda|_2\leq C_\lambda$ and $E|g_{1t}|_2^2<M_f$),

$$
\begin{aligned}
& T^6\mathbb{P}(|\widetilde{H}_T|_2>C_H) \\
\leq\ & T^6\mathbb{P}(|V_T^{-1}|_2>C_v)+T^6\mathbb{P}(\frac{1}{T}\sum_t|g_{1t}|_2^2>2M_f) \\
\leq\ & o(1)+T^6\mathbb{P}(\frac{1}{T}\sum_t(|g_{1t}|_2^2-\mathbb{E}|g_{1t}|_2^2)>M_f)=o(1).
\end{aligned}
$$

Step 4: Proving $T^6\mathbb{P}(a_{1,2}>CN^{-1}\log^c T)=o(1)$ for some $c,C>0$

In step 2, $a_1=|V_T^{-1}|^2\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T|\frac{1}{N}(e_s'e_t-\mathbb{E}e_s'e_t)|^2$. By steps 3 and 4, with probability at least $1-o(T^{-6})$, $|V_T^{-1}|_2<C$. Thus for $c=2c_1^{-1}$,

$$
\begin{aligned}
T^6\mathbb{P}(a_1>CN^{-1}\log^c T) &\leq T^6\mathbb{P}(C\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T|\frac{1}{\sqrt{N}}(e_s'e_t-\mathbb{E}e_s'e_t)|^2>C\log^c T)+o(1) \\
&\leq T^6\mathbb{P}(C\max_{st}|\frac{1}{\sqrt{N}}(e_s'e_t-\mathbb{E}e_s'e_t)|^2>C\log^c T)+o(1) \\
&\leq T^8\max_{st}\mathbb{P}(|\frac{1}{\sqrt{N}}(e_s'e_t-\mathbb{E}e_s'e_t)|>C\log^{c/2}T) \\
&\leq C\exp(11\log T-C_1C^{c_1}\log T)=o(1), \qquad \text{(G.5)}
\end{aligned}
$$

provided that $C_1C^{c_1}>11$. Similarly,

$$
T^6\mathbb{P}(a_2>CN^{-1}\log^c T)\leq o(1)+T^6\max_s\mathbb{P}(|\frac{1}{N}\Lambda'e_s|_2^2>CN^{-1}\log^c T)=o(1). \qquad \text{(G.6)}
$$

Step 5: Prove $T^6\mathbb{P}(\frac{1}{T}\sum_{t=1}^T|\widetilde{f}_{1t}-\widetilde{H}_T'g_{1t}|_2^2>C(\log T)^c(\frac{1}{N}+\frac{1}{T^2}))=o(1)$ for $c=2/c_1$

By (G.4), and steps 3 and 4, there is $C>0$, with probability at least $1-o(T^{-6})$,

$$
\frac{1}{T}\sum_{t=1}^T|\widetilde{f}_{1t}-\widetilde{H}_T'g_{1t}|_2^2\leq C(d_1+...+d_4),
$$

where

$$\begin{aligned}
d_1 &= \frac{1}{T}\sum_{t=1}^{T}|\frac{1}{TN}\sum_{s=1}^{T}g_{1s}(e_s'e_t-\mathbb{E}e_s'e_t)|_2^2,\\
d_2 &= \frac{1}{T}\sum_{t=1}^{T}|\frac{h_t}{\sqrt{N}}|_2^2,\\
d_3 &= \frac{1}{T}\sum_{t=1}^{T}|\frac{1}{TN}\sum_{s=1}^{T}g_{1s}e_s'\Lambda g_{1t}|_2^2,\\
d_4 &= \frac{1}{T}\sum_{t=1}^{T}|\frac{1}{T}\sum_{s=1}^{T}g_{1s}\sigma_{st}|_2^2, \quad \sigma_{st}=\frac{1}{N}\mathbb{E}e_s'e_t.
\end{aligned}$$

The tail probability of d_2 has already been bounded in (G.6):

$$T^6\mathbb{P}(d_2 > N^{-1}C\log^{2/c_1}T)=o(1).$$

For $x=(\log T)^{2/c_1}m$, $y=(\log T)^{2/c_1}m$, $z=(\log T)^{2/c_1}m$ and sufficiently large m,

$$\begin{aligned}
T^6\max_t\mathbb{P}(|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}g_{1s}(e_s'e_t-\mathbb{E}e_s'e_t)|_2 > x^{1/2}) &\le C\exp(10\log T - C_1x^{c_1/2})=o(1),\\
T^6\mathbb{P}(|\frac{1}{TN}\sum_{s=1}^{T}g_{1s}u_s'\Lambda|_2^2 > (NT)^{-1}y) &\le C\exp(10\log T - C_1y^{c_1/2})=o(1),\\
T^6\mathbb{P}(\max_s|g_{1s}|_2^2 > z) &\le \exp(6\log T - C_1z^{c_1/2})=o(1). \qquad \text{(G.7)}
\end{aligned}$$

Note that $\max_t\sum_{s=1}^{T}|\sigma_{st}|\le C_\sigma$ for some $C_\sigma>0$. Therefore,

$$\begin{aligned}
T^6\mathbb{P}(d_1 > (NT)^{-1}x) &\le T^6\mathbb{P}(\frac{1}{T}\sum_{t=1}^{T}|\frac{1}{TN}\sum_{s=1}^{T}g_{1s}(e_s'e_t-\mathbb{E}e_s'e_t)|_2^2 > (NT)^{-1}x)\\
&\le T^6\max_t\mathbb{P}(|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}g_{1s}(e_s'e_t-\mathbb{E}e_s'e_t)|_2 > x^{1/2})=o(1),\\
T^6\mathbb{P}(d_3 > (NT)^{-1}y) &\le T^6\mathbb{P}(|\frac{1}{TN}\sum_{s=1}^{T}g_{1s}e_s'\Lambda|_2^2 > (NT)^{-1}y)+o(1)=o(1),\\
T^6\mathbb{P}(d_4 > T^{-2}C_\sigma^2z) &\le T^6\max_t\mathbb{P}(|\frac{1}{T}\sum_{s=1}^{T}g_{1s}\sigma_{st}|_2^2 > T^{-2}C_\sigma^2z)\\
&\le T^6\max_t\mathbb{P}(\max_s|g_{1s}|^2(\frac{1}{T}\sum_{s=1}^{T}|\sigma_{st}|)^2 > T^{-2}C_\sigma^2z)\\
&\le T^6\max_t\mathbb{P}(\max_s|g_{1s}|^2 > z)=o(1).
\end{aligned}$$

Together, we have, for $c = \log^{2/c_1}$, with probability at least $1 - o(T^{-6})$,

$$\frac{1}{T}\sum_{t=1}^{T} |\widetilde{f}_{1t} - \widetilde{H}_T' g_{1t}|_2^2 \leq Cm_{NT}^2, \text{ where } m_{NT}^2 := (\log T)^c (\frac{1}{N} + \frac{1}{T^2}).$$

Step 6: finishing the proof

We now work with (G.3) $\widetilde{f}_t - \widehat{f}_t = (\sum_{d=1}^{6} A_{t,d}, 0)'$. Write $Q = \frac{1}{T}\sum_{s=1}^{T} |\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s}|_2^2$. Step 5 proved $Q < Cm_{NT}^2$ with probability at least $1 - o(T^{-9})$. In addition,

$$\mathbb{P}(|f_t|_2 > M(\log T)^{1/c_1}) \leq C \exp(-C_f M^{c_1} (\log T)) = CT^{-C_f M^{c_1}} < o(T^{-9})$$

for large enough M.

Now take

$$x = C(\log T)^{1/c_1}, \quad y = C(\log T)^{1/c_1}, \quad w = C(\log T)^{1/c_1},$$

$$z = (\log T)^{1/c_1} w, \quad \widetilde{x} = C(\log T)^{1/c_1}, \quad \widetilde{y} = (\log T)^{1/c_1} \widetilde{x}.$$

Then, we have, for sufficiently large $C > 0$,

$$
\begin{aligned}
T^6\mathbb{P}(|A_{t,1}|_2 > CT^{-1}(\log T)^{1/c_1}) &\le T^6\mathbb{P}(\max_s |g_{1s}|_2 \sum_{s=1}^T |\frac{1}{N}\mathbb{E}e_s' e_t| > C(\log T)^{1/c_1}) + o(1) \\
&\le T^6\mathbb{P}(\max_s |g_{1s}|_2 > C(\log T)^{1/c_1}) + o(1) = o(1), \\
T^6\mathbb{P}(|A_{t,2}|_2 > m_{NT}T^{-1/2}C) &\le T^6\mathbb{P}(|\frac{1}{T}\sum_{s=1}^T(\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s})\frac{1}{N}\mathbb{E}e_s' e_t|_2 > m_{NT}T^{-1/2}C) \\
&\le T^6\mathbb{P}(Q\frac{1}{T}\sum_s |\frac{1}{N}\mathbb{E}e_s' e_t|^2 > m_{NT}^2 T^{-1}C^2) \\
&\le T^6\mathbb{P}(\max_{st} |\frac{1}{N}\mathbb{E}e_s' e_t| \sum_s |\frac{1}{N}\mathbb{E}e_s' e_t| > C^2) + o(1) = o(1), \\
T^6\mathbb{P}(|A_{t,3}|_2 > m_{NT}N^{-1/2}x) &= T^6\mathbb{P}(C|\frac{1}{T}\sum_{s=1}^T(\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s})\frac{1}{N}(e_s' e_t - \mathbb{E}e_s' e_t)| > m_{NT}N^{-1/2}x) + o(1) \\
&\le^{(a)} T^6\mathbb{P}(CQ\frac{1}{T}\sum_{s=1}^T |\frac{1}{N}(e_s' e_t - \mathbb{E}e_s' e_t)|^2 > m_{NT}^2 N^{-1}x^2) + o(1) \\
&\le T^8 \max_{st} \mathbb{P}(|\frac{1}{\sqrt{N}}(e_s' e_t - \mathbb{E}e_s' e_t)| > x) + o(1) =^{(b)} o(1), \\
T^6\mathbb{P}(|A_{t,4}|_2 > (NT)^{-1/2}y) &= T^6\mathbb{P}(C|\frac{1}{\sqrt{TN}}\sum_{s=1}^T g_{1s}(e_s' e_t - \mathbb{E}e_s' e_t)|_2 > y) =^{(c)} o(1) \\
T^6\mathbb{P}(|A_{t,5}|_2 > m_{NT}N^{-1/2}z) &= T^6\mathbb{P}(C|\frac{1}{T}\sum_{s=1}^T(\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s})g_{1t}'\frac{1}{N}\Lambda' e_s|_2 > m_{NT}N^{-1/2}z) + o(1), \\
&\le T^6\mathbb{P}(C|g_{1t}|_2^2\frac{1}{T}\sum_{s=1}^T |\frac{1}{N}\Lambda' e_s|_2^2 > N^{-1}z^2) + o(1) \\
&\le T^7 \max_s \mathbb{P}(C|\frac{1}{\sqrt{N}}\Lambda' e_s|_2 > w) + o(1) =^{(d)} o(1), \\
T^6\mathbb{P}(|A_{t,6}|_2 > (NT)^{-1/2}\widetilde{y}) &= T^6\mathbb{P}(C|\frac{1}{NT}\sum_{s=1}^T g_{1s}g_{1t}'\Lambda' e_s|_2 > (NT)^{-1/2}\widetilde{y}) + o(1) \\
&\le T^6\mathbb{P}(C|\frac{1}{NT}\sum_{s=1}^T g_{1s}e_s'\Lambda|_2 > (NT)^{-1/2}\widetilde{x}) + o(1),
\end{aligned}
$$

where in (a) we used Cauchy-Schwarz; (b) comes from (G.5); (c) and (e) follow from (G.7); (d) is from (G.6). Combined together, $|\widetilde{f}_t - \widehat{f}_t| < C\Delta_f$ with probability at least $1 - o(T^{-9})$,

$$
\begin{aligned}
\Delta_f &= \frac{\log^{1/c_1} T}{T} + \frac{\log^{1/c_1} T + \log^{1/c_1} T \log^{1/c_1} T}{\sqrt{NT}} + m_{NT}(\frac{1}{\sqrt{T}} + \frac{\log^{1/c_1} T}{\sqrt{N}}) \\
&\le 3\frac{\log^{2/c_1} T}{T}.
\end{aligned}
$$

where that last inequality is due to $T = O(N)$.

∎

G.3.2 Defining notation

In the sequel, we show that $(\widehat{\alpha}, \widehat{\gamma})$ defined in Section 4 is asymptotically equivalent to the minimizer of the criterion function that replaces $\widetilde{f}_t$ in $\widetilde{\mathbb{S}}_T(\alpha, \gamma)$ with $\widehat{f}_t$ in the sense that they have an identical asymptotic distribution. Below we introduce various terms in the form of $\widetilde{\cdot}$ and $\widehat{\cdot}$. They indicate that the corresponding terms contain $\widetilde{f}_t$ and $\widehat{f}_t$ in their definitions, respectively.

Let $1_t = 1\{f_t'\gamma_0 > 0\}$ and recall that

$$
\begin{aligned}
& \widetilde{\mathbb{S}}_T(\alpha, \gamma) \\
= \; & \widetilde{\mathbb{S}}_T(\alpha_0, \gamma_0) + \frac{1}{T}\sum_{t=1}^{T}\Big(x_t'(\beta - \beta_0) + x_t'\Big(\delta 1\{\widetilde{f}_t'\gamma > 0\} - \delta_0 1\{\widetilde{f}_t'\gamma_0 > 0\}\Big)\Big)^2 \\
& -\frac{2}{T}\sum_{t=1}^{T}\Big(\varepsilon_t - x_t'\delta_0\Big(1\{\widetilde{f}_t'\gamma_0 > 0\} - 1_t\Big)\Big)\Big(x_t'(\beta - \beta_0) + x_t'\Big(\delta 1\{\widetilde{f}_t'\gamma > 0\} - \delta_0 1\{\widetilde{f}_t'\gamma_0 > 0\}\Big)\Big).
\end{aligned}
$$

And introduce the following decomposition:

$$
\begin{aligned}
\widetilde{\mathbb{S}}_T(\widehat{\alpha}, \widehat{\gamma}) - \widetilde{\mathbb{S}}_T(\alpha_0, \gamma_0) \;=\; & \underbrace{\widetilde{R}_1(\widehat{\alpha}, \widehat{\gamma}) + \widetilde{R}_2(\widehat{\alpha}, \widehat{\gamma}) + \widetilde{R}_3(\widehat{\alpha}, \widehat{\gamma})}_{\widetilde{\mathbb{R}}_T(\widehat{\alpha}, \widehat{\gamma})} \\
& - \underbrace{\Big(\widetilde{\mathbb{C}}_1(\widehat{\alpha}, \widehat{\gamma}) + \widetilde{\mathbb{C}}_2(\widehat{\alpha}, \widehat{\gamma}) - \widetilde{\mathbb{C}}_3(\widehat{\alpha}, \widehat{\gamma}) - \widetilde{\mathbb{C}}_4(\widehat{\alpha}, \widehat{\gamma})\Big)}_{\widetilde{\mathbb{G}}_T(\widehat{\alpha}, \widehat{\gamma}) - \widetilde{\mathbb{G}}_T(\alpha_0, \gamma_0)},
\end{aligned}
$$

where the additional terms are defined in the sequel. Also, note that we suppress the dependence on T to save notational burden as we introduce the more detailed decomposition. Let

$$
\widetilde{Z}_t(\gamma) = (x_t', x_t' 1\{\widetilde{f}_t'\gamma > 0\})', \quad \widehat{Z}_t(\gamma) = (x_t', x_t' 1\{\widehat{f}_t\gamma > 0\})',
$$

$$
\begin{aligned}
\widetilde{\mathbb{R}}_T(\alpha,\gamma) &= \frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{Z}_t(\gamma)'\alpha - \widetilde{Z}_t(\gamma_0)'\alpha_0\right)^2 \\
&= \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{Z}_t(\gamma)'(\alpha-\alpha_0)\right)^2}_{\widetilde{R}_1(\alpha,\gamma)} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(x_t'\delta_0)^2\left|1\left\{\widetilde{f}_t'\gamma>0\right\}-1\left\{\widetilde{f}_t'\gamma_0>0\right\}\right|}_{\widetilde{R}_2(\alpha,\gamma)} \\
&\quad \underbrace{+\frac{2}{T}\sum_{t=1}^{T}x_t'\delta_0\left(1\left\{\widetilde{f}_t'\gamma>0\right\}-1\left\{\widetilde{f}_t'\gamma_0>0\right\}\right)\widetilde{Z}_t(\gamma)'(\alpha-\alpha_0)}_{\widetilde{R}_3(\alpha,\gamma)}, \\
\widetilde{\mathbb{G}}_T(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}\left(\varepsilon_t - x_t'\delta_0\left(1\{\widetilde{f}_t'\gamma_0>0\}-1\{f_t'\gamma_0>0\}\right)\right)\left(\widetilde{Z}_t(\gamma)'\alpha-\widetilde{Z}_t(\gamma_0)'\alpha_0\right).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\widetilde{\mathbb{G}}_T(\alpha,\gamma)-\widetilde{\mathbb{G}}_T(\alpha_0,\gamma_0) &= \frac{2}{T}\sum_{t=1}^{T}\left(\varepsilon_t - x_t'\delta_0\left(1\{\widetilde{f}_t'\gamma_0>0\}-1_t\right)\right)\left(\widetilde{Z}_t(\gamma)'\alpha-\widetilde{Z}_t(\gamma_0)'\alpha_0\right) \\
&= \widetilde{\mathbb{C}}_1(\alpha,\gamma)+\widetilde{\mathbb{C}}_2(\alpha,\gamma)-\widetilde{\mathbb{C}}_3(\alpha,\gamma)-\widetilde{\mathbb{C}}_4(\alpha,\gamma),
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{\mathbb{C}}_1(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\delta\left(1\{\widetilde{f}_t'\gamma>0\}-1\{\widetilde{f}_t'\gamma_0>0\}\right), \\
\widetilde{\mathbb{C}}_2(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}\varepsilon_t\widetilde{Z}_t(\gamma_0)'(\alpha-\alpha_0), \\
\widetilde{\mathbb{C}}_3(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}x_t'\delta_0 x_t'\delta\left(1\{\widetilde{f}_t'\gamma_0>0\}-1_t\right)\left(1\{\widetilde{f}_t'\gamma>0\}-1\{\widetilde{f}_t'\gamma_0>0\}\right), \\
\widetilde{\mathbb{C}}_4(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}x_t'\delta_0\left(1\{\widetilde{f}_t'\gamma_0>0\}-1_t\right)\widetilde{Z}_t(\gamma_0)'(\alpha-\alpha_0).
\end{aligned}
$$

In addition, the following quantities will be used in the proofs to follow.

$$
\begin{aligned}
\widehat{R}_1(\alpha,\gamma) &= \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{Z}_t(\gamma)'(\alpha-\alpha_0)\right)^2, \\
\widehat{R}_2(\alpha,\gamma) &= \frac{1}{T}\sum_{t=1}^{T}(x_t'\delta_0)^2\left|1\{\widehat{f}_t'\gamma>0\}-1\{\widehat{f}_t'\gamma_0>0\}\right|, \\
\widehat{R}_3(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}x_t'\delta_0\left(1\left\{\widehat{f}_t'\gamma>0\right\}-1\left\{\widehat{f}_t'\gamma_0>0\right\}\right)\widehat{Z}_t(\gamma)'(\alpha-\alpha_0),
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\mathbb{C}}_1(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\delta\left(1\{\widehat{f}_t'\gamma>0\}-1\{\widehat{f}_t'\gamma_0>0\}\right),\\
\widehat{\mathbb{C}}_2(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}\varepsilon_t \widehat{Z}_t(\gamma_0)'(\alpha-\alpha_0),\\
\widehat{\mathbb{C}}_3(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}x_t'\delta_0 x_t'\delta\left(1\{\widehat{f}_t'\gamma_0>0\}-1_t\right)\left(1\{\widehat{f}_t'\gamma>0\}-1\{\widehat{f}_t'\gamma_0>0\}\right),\\
\widehat{\mathbb{C}}_4(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^{T}x_t'\delta_0\left(1\{\widehat{f}_t'\gamma_0>0\}-1_t\right)\widehat{Z}_t(\gamma_0)'(\alpha-\alpha_0).
\end{aligned}
$$

G.3.3 Effect of $\widetilde{f}_t-\widehat{f}_t$

Lemma G.1. *Uniformly over α and γ, for Δ_f defined in Proposition G.1,*
(i) For $j=1,...,4$, $\left|\widetilde{\mathbb{C}}_j(\delta,\gamma)-\widehat{\mathbb{C}}_j(\delta,\gamma)\right|\le(T^{-\varphi}+|\alpha-\alpha_0|_2)O_P(\Delta_f+T^{-6})$.
(ii) $|\widetilde{\mathbb{C}}_2(\alpha)|\le O_P(T^{-1/2}+\Delta_f)|\alpha-\alpha_0|_2$.
(iii) $|\widetilde{\mathbb{C}}_4(\alpha)|\le O_P(\Delta_f+N^{-1/2})T^{-\varphi}|\alpha-\alpha_0|_2$.
(iv) For $j=1,2,3$, $|\widetilde{R}_{jT}(\alpha,\gamma)-\widehat{R}_{jT}(\alpha,\gamma)|\le[|\alpha-\alpha_0|_2^2+T^{-2\varphi}]O_P(\Delta_f+T^{-6})$.

A consequence of this lemma is that the first-order asymptotic distribution of $\widehat{\alpha}$ and $\widehat{\gamma}$ can be characterized by the minimizer of $\widehat{\mathbb{S}}_T(\alpha,\gamma)$, which replaces $\widetilde{f}_t$ in the construction of $\widetilde{\mathbb{S}}_T(\alpha,\gamma)$ with $\widehat{f}_t$, since the difference between the two is $T^{-\varphi}O_P(\Delta_f+T^{-6})$, by Proposition G.1. If in addition $T=O(N)$ then it is $T^{-\varphi}O_P(\Delta_f+T^{-6})=o_P\left(T^{-1}\right)$.

Proof. (i) We prove this for $j=1$. The others are similarly shown. Note that

$$
\begin{aligned}
&\sup_{\gamma}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t x_t'[1\{\widetilde{f}_t'\gamma>0\}-1\{\widehat{f}_t'\gamma>0\}]\right|_2\\
\le\;&\sup_{\gamma}\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_t x_t'|_2 1\{\widehat{f}_t'\gamma<0<\widetilde{f}_t'\gamma\}+\sup_{\gamma}\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_t x_t'|_2 1\{\widetilde{f}_t'\gamma<0<\widehat{f}_t'\gamma\}
\end{aligned}
$$

We bound the first term on the right side of the inequality above. The second term follows similarly. As $\sup_\gamma |\gamma|_2 \le C$,

$$
\begin{aligned}
& \sup_\gamma \frac{1}{T}\sum_{t=1}^{T} |\varepsilon_t x_t'|_2 1\{\widehat{f}_t'\gamma < 0 < \widetilde{f}_t'\gamma\} \qquad \text{(G.8)} \\
\le\ & \sup_\gamma \frac{1}{T}\sum_{t=1}^{T} |\varepsilon_t x_t'|_2 1\{-|\widehat{f}_t - \widetilde{f}_t|_2 C < \widehat{f}_t'\gamma < 0\} \\
\le\ & \sup_\gamma \frac{1}{T}\sum_{t=1}^{T} |\varepsilon_t x_t'|_2 1\{\left|\widehat{f}_t'\gamma\right| < C\Delta_f\} + \frac{1}{T}\sum_{t=1}^{T} |\varepsilon_t x_t'|_2 1\{|\widehat{f}_t - \widetilde{f}_t|_2 \ge \Delta_f\} \\
\le\ & \frac{1}{T}\sum_{t=1}^{T} |\varepsilon_t x_t'|_2 1\{\inf_\gamma \left|\widehat{f}_t'\gamma\right| < C\Delta_f\} + O_P(1)\, C\mathbb{P}\{|\widehat{f}_t - \widetilde{f}_t|_2 \ge \Delta_f\} \\
\le\ & O_P(1)\, C\mathbb{P}\left(\inf_\gamma |\widehat{f}_t'\gamma| < C\Delta_f\right) + O_P\left(T^{-6}\right) \\
\le\ & O_P(\Delta_f + T^{-6}),
\end{aligned}
$$

where the first inequality is by the fact that $1\{A\}1\{B\} \le 1\{A\}$ for any events A and B, and the remaining inequalities are by the law of iterated expectations, the rank condition and the moment bound that $\mathbb{E}(|\varepsilon_t x_t|_2 | g_t, h_t) \le C$ a.s. in Assumption 4.1, and Proposition G.1.

(ii) The same proof as in part (i) leads to $\left|\widetilde{\mathbb{C}}_2(\delta,\gamma) - \widehat{\mathbb{C}}_2(\delta,\gamma)\right| \le |\alpha-\alpha_0|_2 O_P(\Delta_f + T^{-6})$.

It suffices to show $|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \widehat{Z}_t(\gamma_0)|_2 \le O_P(\frac{1}{\sqrt{T}})$ due to (i). Then

$$
\begin{aligned}
& |\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \widehat{Z}_t(\gamma_0)|_2 \le O_P(\frac{1}{\sqrt{T}}) + |\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t x_t 1\{\widehat{f}_t'\gamma_0 > 0\}|_2 \\
\le\ & |\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t x_t 1\{(g_t + \frac{h_t}{\sqrt{N}})'\phi_0 > 0\}|_2 + O_P(\frac{1}{\sqrt{T}}) = O_P(\frac{1}{\sqrt{T}}).
\end{aligned}
$$

(iii) The same proof as in part (i) leads to $\left|\widetilde{\mathbb{C}}_4(\delta,\gamma) - \widehat{\mathbb{C}}_4(\delta,\gamma)\right| \le |\alpha-\alpha_0|_2 O_P(n_{NT}+T^{-6})T^{-\varphi}$.

Hence it is sufficient to show that

$$\begin{aligned}
& \frac{1}{T}\sum_t |x_t|_2^2 1\{\widehat{f}_t'\gamma_0 < 0 < f_t'\gamma_0\} \\
\leq\ & \frac{1}{T}\sum_t |x_t|_2^2 1\{0 < f_t'\gamma_0 < |(f_t - \widehat{f}_t)'\gamma_0|\} \leq \frac{1}{T}\sum_t |x_t|_2^2 1\{0 < f_t'\gamma_0 < C\frac{|h_t|_2}{\sqrt{N}}\} \\
\leq\ & O_P(1)\frac{1}{T}\sum_t \mathbb{E}|x_t|_2^2 1\{0 < f_t'\gamma_0 < C\frac{|h_t|_2}{\sqrt{N}}\} \\
\leq\ & O_P(1)\mathbb{E}|x_t|_2^2\mathbb{P}\left(0 < f_t'\gamma_0 < C\frac{|h_t|_2}{\sqrt{N}}\middle| x_t, h_t\right) \\
\leq\ & O_P(1)\mathbb{E}|x_t|_2^2|h_t|_2\frac{1}{\sqrt{N}} = O_P\left(N^{-1/2}\right).
\end{aligned}$$

(iv) Similarly as in (i),

$$\begin{aligned}
& \sup_\gamma |\frac{1}{T}\sum_{t=1}^T x_t\left(1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{\widehat{f}_t'\gamma > 0\right\}\right)\widetilde{Z}_t(\gamma)'| \\
\leq\ & \sup_\gamma \frac{1}{T}\sum_{t=1}^T |x_t|_2^2[1\{\widetilde{f}_t'\gamma < 0 < \widehat{f}_t'\gamma\} + 1\{\widehat{f}_t'\gamma < 0 < \widetilde{f}_t'\gamma\}] \\
\leq\ & \sup_\gamma \frac{2}{T}\sum_{t=1}^T |x_t|_2^2 1\{|\widehat{f}_t'\gamma| < C\Delta_f\} + O_P(T^{-6}) \leq \frac{1}{T}\sum_{t=1}^T |x_t|_2^2 1\{\inf_\gamma |(g_t + \frac{h_t}{\sqrt{N}})'\gamma| < C\Delta_f\} \\
\leq\ & O_P(1)\mathbb{E}|x_t|_2^2\mathbb{P}\left(\inf_\gamma |(g_t + \frac{h_t}{\sqrt{N}})'\gamma| < C\Delta_f \middle| x_t\right) \leq O_P(\Delta_f + T^{-6}).
\end{aligned}$$

Hence uniformly in (α,γ),

$$|\widetilde{R}_3(\alpha,\gamma) - \widehat{R}_3(\alpha,\gamma)| \leq |\alpha - \alpha_0|_2 T^{-\varphi}O_P(\Delta_f + T^{-6})$$

and the cases for $j = 1$ and 2 are similar, so $|\widetilde{R}_1(\alpha,\gamma) - \widehat{R}_1(\alpha,\gamma)| \leq |\alpha - \alpha_0|_2^2 O_P(\Delta_f + T^{-6})$ and $|\widetilde{R}_2(\alpha,\gamma) - \widehat{R}_2(\alpha,\gamma)| \leq T^{-2\varphi}O_P(\Delta_f + T^{-6})$. Together, we have

$$(\Delta_f + T^{-6})[T^{-2\varphi} + |\alpha - \alpha_0|_2^2 + |\alpha - \alpha_0|_2 T^{-\varphi}] \leq 2(\Delta_f + T^{-6})[T^{-2\varphi} + |\alpha - \alpha_0|_2^2].$$

∎

G.3.4 Consistency

The introduced notation $\widehat{R}_i(\alpha,\gamma)$ and $\widehat{\mathbb{C}}_i(\delta,\gamma)$ depend on the random rotation matrix H_T, which is inconvenient to carry throughout the study of consistency and rates of convergence. On the other hand, with $\breve{g}_t := g_t + \frac{1}{\sqrt{N}}h_t$, note that for any γ and $\phi = H_T\gamma$, we have

$\widehat{f}_t'\gamma = \breve{g}_t'\phi$, which is in fact independent of H_T. It is therefore more convenient to work with functions with respect to ϕ. Hence we introduce the following functions of reparametrization:

$$
\begin{aligned}
\breve{\mathbb{Z}}_t(\phi) &= (x_t', x_t' 1\{\breve{g}_t'\phi > 0\})', \\
\mathbb{Z}_t(\phi) &= (x_t', x_t' 1\{g_t'\phi > 0\})', \\
\mathbb{R}(\alpha,\phi) &= \mathbb{E}[(\alpha-\alpha_0)'\mathbb{Z}_t(\phi)]^2, \\
\mathbb{R}_2(\phi) &= \widehat{R}_2\left(\alpha, H_T^{-1}\phi\right) = \frac{1}{T}\sum_{t=1}^T \left(x_t'\delta_0\right)^2 |1\{\breve{g}_t'\phi > 0\} - 1\{\breve{g}_t'\phi_0 > 0\}|, \\
\mathbb{R}_3(\alpha,\phi) &= \widehat{R}_3\left(\alpha, H_T^{-1}\phi\right) = \frac{2}{T}\sum_{t=1}^T x_t'\delta_0 \left(1\left\{\breve{g}_t'\phi > 0\right\} - 1\left\{\breve{g}_t'\phi_0 > 0\right\}\right) \breve{\mathbb{Z}}_t(\phi)'(\alpha-\alpha_0), \\
\mathbb{C}_1(\delta,\phi) &= \widehat{\mathbb{C}}_1\left(\delta, H_T^{-1}\phi\right) = \frac{2}{T}\sum_{t=1}^T \varepsilon_t x_t'\delta \left(1\{\breve{g}_t'\phi > 0\} - 1\{\breve{g}_t'\phi_0 > 0\}\right), \\
\mathbb{C}_3(\delta,\phi) &= \widehat{\mathbb{C}}_3\left(\delta, H_T^{-1}\phi\right) = \frac{2}{T}\sum_{t=1}^T x_t'\delta_0 x_t'\delta \left(1\{\breve{g}_t'\phi_0 > 0\} - 1_t\right)\left(1\{\breve{g}_t'\phi > 0\} - 1\{\breve{g}_t'\phi_0 > 0\}\right).
\end{aligned}
$$

Lemma G.2. *Uniformly in (α,ϕ), for an arbitrarily small $\eta > 0$,*
(i) $\sup_\phi |\widehat{R}_1(\alpha, H_T^{-1}\phi) - \mathbb{R}(\alpha,\phi)| = o_P(1)|\alpha-\alpha_0|_2^2$,
(ii) $|\mathbb{R}_3(\alpha,\phi)| \le \left(O_P\left(T^{-1}\right) + CT^{-\varphi}|\phi-\phi_0|_2\right)|\alpha-\alpha_0|_2$.
(iii) $|\mathbb{C}_1(\delta,\phi) - \mathbb{C}_1(\delta_0,\phi)| \le \left(O_P\left(T^{-1}\right) + \eta T^{-2\varphi}|\phi-\phi_0|\right) T^{\varphi}|\delta-\delta_0|_2$
(iv) $|\mathbb{C}_3(\delta,\phi) - \mathbb{C}_3(\delta_0,\phi)| \le T^{-\varphi}|\delta-\delta_0|_2 O_P\left(N^{-1/2}\right)$.

Proof. (i) First, note that by uniform law of large numbers, for a sufficiently large $C > 0$,

$$
\sup_\phi \left| \frac{1}{T}\sum_{t=1}^T \breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)' - \mathbb{E}\breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)' \right| = o_P(1).
$$

In addition, $\left|\mathbb{E}\mathbb{Z}_t(\phi)\mathbb{Z}_t(\phi)' - \mathbb{E}\breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)'\right| = o_P(1)$. Also, $\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(H_T^{-1}\phi)\widehat{Z}_t(H_T^{-1}\phi)' = \frac{1}{T}\sum_{t=1}^T \breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)'$. Hence

$$
\begin{aligned}
& \sup_\phi |\widehat{R}_1(\alpha, H_T^{-1}\phi) - \mathbb{R}(\alpha,\phi)| \\
\le\ & |\alpha-\alpha_0|_2^2 \sup_\phi \left| \frac{1}{T}\sum_{t=1}^T \breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)' - \mathbb{E}\breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)' \right| \\
& + |\alpha-\alpha_0|_2^2 \sup_\phi \left| \mathbb{E}\mathbb{Z}_t(\phi)\mathbb{Z}_t(\phi)' - \mathbb{E}\breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)' \right| \\
=\ & o_P(1)|\alpha-\alpha_0|_2^2.
\end{aligned}
$$

(ii) By Lemma I.2, uniformly in ϕ

$$\begin{aligned}
|\mathbb{R}_3(\alpha,\phi)| &= |\frac{2}{T}\sum_{t=1}^{T} x_t'\delta_0\left(1\left\{\breve{g}_t'\phi>0\right\}-1\left\{\breve{g}_t'\phi_0>0\right\}\right)\breve{\mathbb{Z}}_t(\phi)'(\alpha-\alpha_0)| \\
&\le C|\alpha-\alpha_0|_2\frac{1}{T^{1+\varphi}}\sum_{t=1}^{T}|x_t|_2^2\left|1\left\{\breve{g}'\phi>0\right\}-1\left\{\breve{g}'\phi_0>0\right\}\right| \\
&\le C|\alpha-\alpha_0|_2\left[O_P(T^{-1})+T^{-2\varphi}|\phi-\phi_0|\right] \\
&\quad +C|\alpha-\alpha_0|_2T^{-\varphi}\mathbb{E}|x_t|_2^2\left|1\left\{\breve{g}'\phi>0\right\}-1\left\{\breve{g}'\phi_0>0\right\}\right| \\
&\le C|\alpha-\alpha_0|_2\left[O_P(T^{-1})+T^{-2\varphi}|\phi-\phi_0|\right].
\end{aligned}$$

(iii) Due to Lemma I.2 and Hölder inequality, for an arbitrarily small $\eta>0$,

$$|\mathbb{C}_1(\delta,\phi)-\mathbb{C}_1(\delta_0,\phi)| \le \left|\frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t\left(1\{\breve{g}_t'\phi>0\}-1\{\breve{g}_t'\phi_0>0\}\right)\right||\delta-\delta_0|_2$$

$$\begin{aligned}
&= \left|\frac{2}{T^{1+\varphi}}\sum_{t=1}^{T}\varepsilon_t x_t\left(1\left\{\breve{g}'\phi_0\le 0<\breve{g}'\phi\right\}-1\left\{\breve{g}'\phi\le 0<\breve{g}'\phi_0\right\}\right)\right| \\
&\quad T^{\varphi}|\delta-\delta_0|_2 \\
&\le \left(O_P\left(T^{-1}\right)+\eta T^{-2\varphi}\left|\phi-\phi_0\right|\right)T^{\varphi}\left|\delta-\delta_0\right|_2.
\end{aligned}$$

(iv) Uniformly in ϕ,

$$\begin{aligned}
\left|\mathbb{C}_3\left(\delta_0,\phi\right)-\mathbb{C}_3\left(\delta,\phi\right)\right| &\le \frac{2}{T}\sum_{t=1}^{T}|x_t|_2^2\left|1\{\breve{g}_t'\phi_0>0\}-1\{g_t'\phi_0>0\}\right||\delta-\delta_0|_2T^{-\varphi} \\
&\le T^{-\varphi}\left|\delta-\delta_0\right|_2 O_P\left(N^{-1/2}\right),
\end{aligned}$$

since the modulus of the difference between two indicators is less than equal to 1. ∎

Proposition G.2.

$$|\widehat{\alpha}-\alpha_0|_2=o_P(1), \quad |\widehat{\phi}-\phi_0|_2=o_P(1).$$

Since $H_T^{-1}=O_P(1)$, this proposition implies that $\widehat{\gamma}-\gamma_0=H_T^{-1}(\widehat{\phi}-\phi_0)+o_P(1)=o_P(1)$ as well.

Proof. We begin with showing the consistency of $\widehat{\gamma}$. Let $\widetilde{P}(\gamma)$ and $\widehat{P}(\gamma)$ respectively be the orthogonal projection matrices on $\widetilde{Z}_t(\gamma)$ and $\widehat{Z}_t(\gamma)$. Then

$$\begin{aligned}
\widetilde{\mathbb{S}}_T(\gamma) &= \widetilde{\mathbb{S}}_T\left(\widehat{\alpha}(\gamma),\gamma\right)=\frac{1}{T}Y'\left(I-\widetilde{P}(\gamma)\right)Y \\
&= \frac{1}{T}\left(e'\left(I-\widetilde{P}(\gamma)\right)e+2\delta_0'X_0\left(I-\widetilde{P}(\gamma)\right)e+\delta_0'X_0'\left(I-\widetilde{P}(\gamma)\right)X_0\delta_0\right),
\end{aligned}$$

where e, Y, and X_0 are the matrices stacking ε_t's, y_t's and $x_t' 1_t$'s, respectively.

Let $\widetilde{\gamma}$ be an estimator such that

$$\widetilde{\mathbb{S}}_T(\widetilde{\gamma}) \leq \widetilde{\mathbb{S}}_T(\gamma_0) + o_P\left(T^{-2\varphi}\right). \tag{G.9}$$

Then, $\widehat{\gamma}$ satisfies this as it is a minimizer. Furthermore,

$$\begin{aligned} 0 &\geq T^{2\varphi}\left(\widetilde{\mathbb{S}}_T(\widetilde{\gamma}) - \widetilde{\mathbb{S}}_T(\gamma_0)\right) - o_P(1) \\ &= \frac{T^{2\varphi}}{T}\left(e'\left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right)e + 2\delta_0' X_0\left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right)e + \delta_0' X_0'\left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right)X_0\delta_0\right). \end{aligned} \tag{G.10}$$

For the first term in (G.10), recall $\breve{g}_t = g_t + h_t N^{-1/2}$ and note that by Lemma G.1, Lemma G.2 and ULLN lead to uniformly in γ, and $\phi = H_T\gamma$, (recall $\mathbb{Z}_t(\phi) = Z_t(\gamma)$)

$$\begin{aligned} &\frac{1}{T}\widetilde{Z}(\gamma)'\widetilde{Z}(\gamma) = \frac{1}{T}\widehat{Z}(\gamma)'\widehat{Z}(\gamma) + o_P(1) = T^{-1}\sum_{t=1}^{T} Z_t(\gamma) Z_t(\gamma)' + o_P(1) \\ = \; & T^{-1}\sum_{t=1}^{T}\mathbb{Z}_t(\phi)\mathbb{Z}_t(\phi)' + o_P(1) = \mathbb{E}\mathbb{Z}_t(\phi)\mathbb{Z}_t(\phi)' + o_P(1). \end{aligned}$$

Then the rank condition for $\mathbb{E}\mathbb{Z}_t(\phi)\mathbb{Z}_t(\phi)'$ in Assumption 4.1 implies that $\sup_\gamma[\frac{1}{T}\widetilde{Z}(\gamma)'\widetilde{Z}(\gamma)]^{-1} = O_P(1)$. Also,

$$\sup_\gamma |\frac{1}{T}\widetilde{Z}(\gamma)'e|_2 \leq \sup_\gamma |\frac{1}{T}\widehat{Z}(\gamma)'e|_2 + O_P(\Delta_f + T^{-6}) = O_P(\frac{1}{\sqrt{T}}),$$

by Lemma G.1 and an FCLT for VC classes in Arcones and Yu (1994). So

$$\begin{aligned} &|\frac{1}{T}e'\left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right)e| \leq 2\sup_\gamma \frac{1}{T}e'\widetilde{P}(\gamma)e \leq 2\frac{1}{T}\sup_\gamma |[\widetilde{Z}(\gamma)'\widetilde{Z}(\gamma)]^{-1}|_2^2 |\widetilde{Z}(\gamma)'e|_2^2 \\ \leq \; & 2\sup_\gamma[\frac{1}{T}\widetilde{Z}(\gamma)'\widetilde{Z}(\gamma)]^{-1}\sup_\gamma|\frac{1}{T}\widetilde{Z}(\gamma)'e|_2^2 = O_P(T^{-1}). \end{aligned}$$

So $\frac{T^{2\varphi}}{T}e'\left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right)e = o_P(1)$. For the second term in (G.10),

$$\begin{aligned} &\frac{T^{2\varphi}}{T}\delta_0' X_0\left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right)e \leq O_P(T^\varphi)\sup_\gamma|\frac{1}{T}X_0\widetilde{P}(\gamma)e|_2 \\ \leq \; & O_P(T^\varphi)\sup_\gamma|\frac{1}{T}\sum_t X_t\varepsilon_t 1\{\widetilde{f}_t'\gamma > 0\}| \\ = \; & O_P(T^\varphi)\sup_\gamma|\frac{1}{T}\sum_t X_t\varepsilon_t 1\{\widehat{f}_t'\gamma > 0\}| + O_P(T^\varphi)(\Delta_f + T^{-6}) \\ = \; & o_P(1), \end{aligned}$$

due to Lemma G.1 and FCLT. Applying the same reasoning for the third term in (G.10) and recalling that $P(\gamma_0)X_0 = X_0$,

$$\frac{T^{2\varphi}}{T}\delta_0' X_0' \left(\widetilde{P}(\gamma_0) - \widetilde{P}(\widetilde{\gamma})\right) X_0\delta_0 = o_P(1) + \mathbb{E}(d_0' x_t)^2 1_t - A(\widetilde{\phi}),$$

where $A(\widetilde{\phi}) = \mathbb{E}d_0' x_t 1_t \mathbb{Z}_t\left(\widetilde{\phi}\right)' \left(\mathbb{E}\mathbb{Z}_t\left(\widetilde{\phi}\right)\mathbb{Z}_t\left(\widetilde{\phi}\right)'\right)^{-1} \mathbb{E}\mathbb{Z}_t\left(\widetilde{\phi}\right) 1_t x_t' d_0$. The remaining proof for $\widetilde{\phi} \xrightarrow{P} \phi_0$ is the same as the known factor case.

Turning to $\widehat{\alpha}$, recall

$$\widetilde{\mathbb{R}}_T\left(\alpha, H_T^{-1}\phi\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{Z}_t\left(H_T^{-1}\phi\right)'\alpha - \widetilde{Z}_t\left(H_T^{-1}\phi_0\right)'\alpha_0\right)^2.$$

Write

$$\begin{aligned}
\mathbb{R}(\alpha,\phi) &:= \mathbb{E}\left(\breve{\mathbb{Z}}_t(\phi)'\alpha - \breve{\mathbb{Z}}_t(\phi_0)'\alpha_0\right)^2 \\
\mathbb{R}^0(\alpha,\phi) &:= \mathbb{E}\left(\mathbb{Z}_t(\phi)'\alpha - \mathbb{Z}_t(\phi_0)'\alpha_0\right)^2.
\end{aligned}$$

We have

$$\begin{aligned}
&\sup_{\alpha,\phi}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{Z}_t\left(H_T^{-1}\phi\right)'\alpha - \widetilde{Z}_t\left(H_T^{-1}\phi_0\right)'\alpha_0\right)^2 - \left(\widehat{Z}_t\left(H_T^{-1}\phi\right)'\alpha - \widehat{Z}_t\left(H_T^{-1}\phi_0\right)'\alpha_0\right)^2\right| \\
\leq\ & \sup_{\phi}\left(\frac{1}{T}\sum_{t=1}^{T}|x_t|_2^2 1\{|\breve{g}_t'\phi| < |\widehat{f}_t - \widetilde{f}_t|_2 C\}\right)^{1/2} \\
\leq\ & \left(\frac{1}{T}\sum_{t=1}^{T}|x_t|_2^2 1\{\inf_{\phi}|\breve{g}_t'\phi| < |\widehat{f}_t - \widetilde{f}_t|_2 C\}\right)^{1/2} \\
\leq\ & \left(\frac{1}{T}\sum_{t=1}^{T}|x_t|_2^2 1\{\inf_{\phi}|\breve{g}_t'\phi| < \Delta_f C\} + \frac{1}{T}\sum_{t=1}^{T}|x_t|_2^2 1\{|\widehat{f}_t - \widetilde{f}_t| > \Delta_f,\ \text{or}\ |H_T| > C\}\right)^{1/2} \\
=\ & o_P(1).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\sup_{\alpha,\phi}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{Z}_t\left(H_T^{-1}\phi\right)'\alpha - \widehat{Z}_t\left(H_T^{-1}\phi_0\right)'\alpha_0\right)^2 - \mathbb{R}(\alpha,\phi)\right| \\
=\ & \sup_{\alpha,\phi}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{Z}_t(\phi)'\alpha - \breve{\mathbb{Z}}_t(\phi_0)'\alpha_0\right)^2 - \mathbb{R}(\alpha,\phi)\right| = o_P(1),
\end{aligned}$$

by uniform law of large numbers. Also,

$$\sup_{\alpha,\phi} |\mathbb{R}(\alpha, \phi) - \mathbb{R}^0(\alpha, \phi)| \leq \left(\mathbb{E}|x_t|_2^2 1\{\inf_{\phi} |g_t'\phi| < C|h_t|_2 N^{-1/2}\}\right)^{1/2} = o(1).$$

Hence $\sup_{\alpha,\phi} \left|\widetilde{\mathbb{R}}_T\left(\alpha, H_T^{-1}\phi\right) - \mathbb{R}^0(\alpha, \phi)\right| \leq o_P(1)$.

Next, we turn to the $\widehat{\phi}$. Recall that $\widehat{\alpha}$ and $\widehat{\gamma}$ are minimizers of $\widetilde{\mathbb{S}}_T$ and thus

$$0 \geq \widetilde{\mathbb{S}}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) - \widetilde{\mathbb{S}}_T\left(\alpha_0, \gamma_0\right) = \widetilde{\mathbb{R}}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) - \widetilde{\mathbb{G}}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) + \widetilde{\mathbb{G}}_T\left(\alpha_0, \gamma_0\right).$$

Since $\widehat{\phi} := H_T\widehat{\gamma}$, Lemma G.1, G.2, and the fact that $\mathbb{C}_i\left(\delta, \widehat{\phi}\right) = \widehat{\mathbb{C}}_i\left(\delta, \widehat{\gamma}\right)$, $i = 1, 3$ imply that

$$\begin{aligned} |\mathbb{R}^0(\widehat{\alpha}, \widehat{\phi})| &\leq \widetilde{\mathbb{R}}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) + \sup_{\alpha,\phi}\left|\widetilde{\mathbb{R}}_T\left(\alpha, H_T^{-1}\phi\right) - \mathbb{R}^0(\alpha, \phi)\right| \\ &\leq o_P(1) + \widetilde{\mathbb{G}}_T\left(\widehat{\alpha}, \widehat{\gamma}\right) + \widetilde{\mathbb{G}}_T\left(\alpha_0, \gamma_0\right) \\ &\leq o_P(1) + |\widetilde{\mathbb{C}}_1(\widehat{\delta}, \widehat{\gamma})| + |\widetilde{\mathbb{C}}_2(\widehat{\alpha})| + |\widetilde{\mathbb{C}}_3(\widehat{\delta}, \widehat{\gamma})| + |\widetilde{\mathbb{C}}_4(\widehat{\alpha})| \\ &\leq o_P(1) + |\widehat{\mathbb{C}}_1(\delta_0, \widehat{\gamma})| + |\widehat{\mathbb{C}}_3(\delta_0, \widehat{\gamma})| = o_P(1). \end{aligned}$$

By the identification theorem, $\mathbb{R}^0(\alpha, \phi)$ has a unique minimum at (α_0, ϕ_0). Then the continuity of $\mathbb{R}^0$ implies $\widehat{\alpha} \xrightarrow{P} \alpha_0$ and $\widehat{\phi} \xrightarrow{P} \phi_0$ by the argmax continuous mapping theorem (see e.g. van der Vaart and Wellner, 1996, p.286). ∎

G.4 Rate of convergence for $\widehat{\phi}$ (Proof of Theorem 4.1)

Here, we prove Theorem 4.1. Let

$$\begin{aligned} \mathbb{G}_1(\phi) &:= \mathbb{E}\mathbb{R}_2(\phi) + \mathbb{E}\mathbb{C}_3(\delta_0, \phi) \\ \mathbb{G}_2(\phi) &:= |\mathbb{R}_2(\phi) + \mathbb{C}_3(\delta_0, \phi) - (\mathbb{E}\mathbb{R}_2(\phi) + \mathbb{E}\mathbb{C}_3(\delta_0, \phi))|. \end{aligned} \tag{G.11}$$

Recall that $\mathbb{R}(\alpha, \phi) = \mathbb{E}[(\alpha - \alpha_0)'\mathbb{Z}_t(\phi)]^2$.

Lemma G.3. *Uniformly in α, ϕ, for any $\epsilon > 0$, there is $C > 0$ that is independent of ϵ, and C_ϵ that depends on ϵ, so that $|\mathbb{R}(\alpha, \phi) - \mathbb{R}(\alpha, \phi_0)| \leq C|\alpha - \alpha_0|_2^2[C_\epsilon|\phi - \phi_0|_2 + \epsilon]^{1/2}$. Hence $|\mathbb{R}(\alpha, \widehat{\phi}) - \mathbb{R}(\alpha, \phi_0)| = o_P(1)|\alpha - \alpha_0|_2^2$.*

Proof. For any $\epsilon > 0$, there is C_1, so that $\mathbb{P}(|g_t|_2 > C_1) < \epsilon$. Note that for any deterministic ϕ,

$$\begin{aligned} &|\mathbb{R}(\alpha, \phi) - \mathbb{R}(\alpha, \phi_0)| \leq |\alpha - \alpha_0|_2^2 \mathbb{E}|x_t|_2^2 1\{|g_t'\phi_0| < |g_t|_2|\phi - \phi_0|_2\} \\ \leq\ & |\alpha - \alpha_0|_2^2 \mathbb{P}^{1/2}(|g_t'\phi_0| < |g_t|_2|\phi - \phi_0|_2)(\mathbb{E}|x_t|_2^4)^{1/2} \\ \leq\ & C|\alpha - \alpha_0|_2^2[\mathbb{P}(|g_t'\phi_0| < C_\epsilon|\phi - \phi_0|_2) + \mathbb{P}(|g_t|_2 > C_1)]^{1/2} \end{aligned}$$

$$\leq \quad C|\alpha-\alpha_0|_2^2[C_\epsilon|\phi-\phi_0|_2+\epsilon]^{1/2}.$$

Now let $\phi=\widehat{\phi}$, and the consistency implies $|\widehat{\phi}-\phi_0|_2=o_P(1)$. Thus

$$|\mathbb{R}(\alpha,\phi)-\mathbb{R}(\alpha,\phi_0)|\leq C|\alpha-\alpha_0|_2^2[C_\epsilon o_P(1)+\epsilon]^{1/2}.$$

Since $\epsilon>0$ is arbitrary, we have the desired result. ∎

Lemma G.4. *For an arbitrarily small $\eta>0$, uniformly in ϕ,*

$$|\mathbb{G}_2(\phi))|\leq b_{NT}T^{-\varphi},\quad |\mathbb{C}_1\left(\delta_0,\phi\right)|\leq b_{NT}.$$

If in addition, $\sqrt{N}=O(T^{1-2\varphi})$, then

$$|\mathbb{G}_2(\phi)|\leq a_{NT}T^{-\varphi},\quad |\mathbb{C}_1\left(\delta_0,\phi\right)|\leq a_{NT}.$$

where

$$\begin{aligned}
a_{NT} &= T^{-2\varphi}O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right)+T^{-2\varphi}\eta\left|\phi-\phi_0\right|_2^2\sqrt{N}\\
b_{NT} &= O_P(\frac{1}{T})+\eta T^{-2\varphi}\left|\phi-\phi_0\right|_2.
\end{aligned}$$

Proof. Let $z_t=T^{2\varphi}2(x_t'\delta_0)^2\left(1\{\breve{g}_t'\phi_0>0\}-1(g_t'\phi_0>0)\right)$. By Lemma I.2, we have the following bound:

$$\begin{aligned}
|\mathbb{C}_3(\delta_0,\phi)-\mathbb{E}\mathbb{C}_3(\delta_0,\phi)| &= T^{-\varphi}|\frac{1}{T^{1+\varphi}}\sum_{t=1}^{T}[z_t\left(1\{\breve{g}_t'\phi>0\}-1\{\breve{g}_t'\phi_0>0\}\right)\\
&\quad -\mathbb{E}z_t\left(1\{\breve{g}_t'\phi>0\}-1\{\breve{g}_t'\phi_0>0\}\right)]|\\
&\leq O_P(\frac{1}{T^{1+\varphi}})+\eta T^{-3\varphi}\left|\phi-\phi_0\right|_2.
\end{aligned}$$

In addition, by Lemma I.3, when $\sqrt{N}=O(T^{1-2\varphi})$ we have the other upper bound:

$$\begin{aligned}
|\mathbb{C}_3(\delta_0,\phi)-\mathbb{E}\mathbb{C}_3(\delta_0,\phi)| &= T^{-3\varphi}|\frac{1}{T^{1-\varphi}}\sum_{t=1}^{T}[z_t\left(1\{\breve{g}_t'\phi>0\}-1\{\breve{g}_t'\phi_0>0\}\right)\\
&\quad -\mathbb{E}z_t\left(1\{\breve{g}_t'\phi>0\}-1\{\breve{g}_t'\phi_0>0\}\right)]|\\
&\leq T^{-3\varphi}O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right)+T^{-3\varphi}\eta\left|\phi-\phi_0\right|_2^2\sqrt{N}
\end{aligned}$$

Similarly, the same upper bound applies to $|\mathbb{R}_2(\phi)-\mathbb{E}\mathbb{R}_2(\phi)|$.

Furthermore, note that for any $\eta > 0$

$$\begin{aligned}\mathbb{C}_1(\delta_0,\phi) &\leq \Big|\frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\delta_0\big(1\{\breve{g}_t'\phi_0 \leq 0 < \breve{g}_t'\phi\} - 1\{\breve{g}_t'\phi \leq 0 < \breve{g}_t'\phi_0\}\big)\Big| \\ &\leq O_P\left(T^{-1}\right) + \eta T^{-2\varphi}\left|\phi-\phi_0\right|_2\end{aligned}$$

due to Lemma I.2 and that when $\sqrt{N} = O(T^{1-2\varphi})$

$$\mathbb{C}_1(\delta_0,\phi) \leq T^{-2\varphi}O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right) + \eta T^{-2\varphi}\sqrt{N}\left|\phi-\phi_0\right|^2, \tag{G.12}$$

due to Lemma I.3. ∎

Lemma G.5 below holds regardless of whether $N^{1/2} < T^{1-2\varphi}$ or not, but is crude when $N^{1/2} = o(T^{1-2\varphi})$. When $N^{1/2} = o(T^{1-2\varphi})$, a sharper bound is given in Lemma G.6.

Lemma G.5. *Suppose the conditional density of $f_t'\gamma_0$ given (x_t, h_t) is bounded away from above almost surely. Then there is a constant $C, c > 0$ that do not depend on ϕ,*

$$\mathbb{G}_1(\phi) \geq cT^{-2\varphi}|\phi-\phi_0|_2 - \frac{C}{\sqrt{N}T^{2\varphi}}.$$

Proof. First,

$$|\mathbb{E}\mathbb{C}_3(\delta_0,\phi)| \leq \mathbb{E}(x_t'\delta_0)^2\left|1\{\breve{g}_t'\phi_0 > 0\} - 1\{g_t'\phi_0 > 0\}\right| \leq CT^{-2\varphi}\frac{1}{\sqrt{N}}.$$

Next, we lower bound $\mathbb{E}\mathbb{R}_2(\phi) = \mathbb{E}\left(x_t'\delta_0\right)^2\left|1\{\breve{g}_t'\phi > 0\} - 1\{\breve{g}_t'\phi_0 > 0\}\right|$. The proof is similar to Step 1 of Proof of Lemma F.3]. We show that there exists a constant $c > 0$ and a neighborhood of ϕ_0 such that for all ϕ in the neighborhood

$$G(\gamma) = \mathbb{E}\left|1\{\breve{g}_t'\phi > 0\} - 1\{\breve{g}_t'\phi_0 > 0\}\right| \geq c\left|\phi-\phi_0\right|_2.$$

Note that the first element of $(\gamma-\gamma_0)$ is zero due to the normalization. Then,

$$G(\gamma) = \mathbb{P}\left\{-\widehat{f}_{2t}'(\gamma_2-\gamma_{20}) \leq \breve{g}_t'\phi_0 < 0\right\} + \mathbb{P}\left\{0 < \breve{g}_t'\phi_0 \leq -\widehat{f}_{2t}'(\gamma_2-\gamma_{20})\right\}.$$

Since the conditional density of $\breve{g}_t'\phi_0$ given $\widehat{f}_{2t}$ is bounded away from zero and continuous in a sufficiently small open neighborhood ϵ of zero, we can find $c_1 > 0$ so that

$$\mathbb{P}\left\{-\widehat{f}_{2t}'(\gamma_2-\gamma_{20}) \leq \breve{g}_t'\phi_0 < 0\right\} \geq c_1\mathbb{E}\left(\widehat{f}_{2t}'(\gamma_2-\gamma_{20})1\left\{\widehat{f}_{2t}'(\gamma_2-\gamma_{20}) > 0\right\}1\left\{\left|\widehat{f}_{2t}'\right| \leq M\right\}\right),$$

where M satisfies that $\left|\gamma-\gamma_0\right|_2 M < \epsilon$. This is always feasible because we can make $\left|\gamma-\gamma_0\right|_2$

as small as necessary due to the consistency of $\widehat{\gamma}$. Similarly,

$$\mathbb{P}\left\{0 < \breve{g}_t' \phi_0 \le -\widehat{f}_{2t}' (\gamma_2 - \gamma_{20})\right\} \ge c_1 \mathbb{E}\left(-\widehat{f}_{2t}' (\gamma_2 - \gamma_{20})\, 1\left\{\widehat{f}_{2t}' (\gamma_2 - \gamma_{20}) < 0\right\} 1\left\{\left|\widehat{f}_{2t}'\right| \le M\right\}\right).$$

Thus,

$$G(\gamma) \ge c_1 \mathbb{E}\left(\left|\widehat{f}_{2t}' (\gamma_2 - \gamma_{20})\right| 1\left\{\left|\widehat{f}_{2t}\right| \le M\right\}\right) \ge c_2 |\gamma - \gamma_0|_2$$

for some $c_2 > 0$ because

$$\inf_{|r|=1} \mathbb{E}\left(\left|\widehat{f}_{2t}' r\right| 1\left\{\left|\widehat{f}_{2t}\right| \le M\right\}\right) > 0$$

for some $M < \infty$. The last inequality $\inf_{|r|=1} \mathbb{E}\left(\left|\widehat{f}_{2t}' r\right| 1\left\{\left|\widehat{f}_{2t}\right| \le M\right\}\right) > 0$ follows since

$$\begin{aligned}
&\inf_{|r|=1} \mathbb{E}\left(\left|\widehat{f}_{2t}' r\right| 1\left\{\left|\widehat{f}_{2t}\right| \le M\right\}\right) \\
&\ge \inf_{|r|=1} \mathbb{E}\left(\left|f_{2t}' r\right| 1\left\{|f_{2t}| \le M\right\}\right) - \mathbb{E}|\widehat{f}_t - f_t|_2 - \mathbb{E}|f_t|_2 1\{M - \frac{|h_t|_2}{\sqrt{N}} < |f_t|_2 < M + \frac{|h_t|_2}{\sqrt{N}}\} \\
&\ge c - O(N^{-1/8}) - \mathbb{E}|f_t|_2 1\{M - \frac{|h_t|_2}{\sqrt{N}} < |f_t|_2 < M + \frac{|h_t|_2}{\sqrt{N}}\} 1\{|h_t|_2 < MN^{1/4}\} \\
&\ge c/2 - c\left[\sup_{|f|<2M, h_t} p_{f_{2t}|h_t}(f) \mathbb{E}\mu\left(f \in \mathbb{R}^{\dim(f_{2t})} : M - \frac{|h_t|_2}{\sqrt{N}} < |f|_2 < M + \frac{|h_t|_2}{\sqrt{N}}\right) 1\{|h_t|_2 < MN^{1/4}\}\right]^{1/2} \\
&\ge c/2 - c\left[\mathbb{E}\left((M + \frac{|h_t|_2}{\sqrt{N}})^{\dim(f_{2t})} - (M - \frac{|h_t|_2}{\sqrt{N}})^{\dim(f_{2t})}\right) 1\{|h_t|_2 < MN^{1/4}\}\right]^{1/2} \ge c/4.
\end{aligned}$$

where $\mu(A)$ denotes the Lebesgue measure of the set A; here A is the difference of two balls in $\mathbb{R}^{\dim(f_{2t})}$. Here the second inequality follows from: $\mathbb{E}|\widehat{f}_t - f_t|_2 = O(N^{-1/2})$, and write $a_t := |f_t|_2 1\{M - \frac{|h_t|_2}{\sqrt{N}} < |f_t|_2 < M + \frac{|h_t|_2}{\sqrt{N}}\}$.

$$\begin{aligned}
\mathbb{E}a_t &\le \mathbb{E}a_t 1\{|h_t|_2 < MN^{1/4}\} + (\mathbb{E}a_t^2)^{1/2} \mathbb{P}(|h_t|_2 > MN^{1/4})^{1/2} \\
&\le \mathbb{E}a_t 1\{|h_t|_2 < MN^{1/4}\} + (\mathbb{E}|f_t|_2^2)^{1/2} (\frac{\mathbb{E}|h_t|_2}{MN^{1/4}})^{1/2} \\
&\le \mathbb{E}a_t 1\{|h_t|_2 < MN^{1/4}\} + O(N^{-1/8}).
\end{aligned}$$

∎

Proposition G.3 (Preliminary Rate of convergence)**.** *Suppose $T^{2\varphi} \log^\kappa T = O(N)$ for any $\kappa > 0$. For $\widehat{\phi} = H_T \widehat{\gamma}$,*

$$|\widehat{\alpha} - \alpha_0|_2 = O_P(T^{-1/2} + N^{-1/4} T^{-\varphi}), \quad |\widehat{\phi} - \phi_0|_2 = O_P(T^{-(1-2\varphi)} + N^{-1/2}).$$

Remark When $T^{1-2\varphi} = O(\sqrt{N})$, this rate becomes

$$|\widehat{\alpha} - \alpha_0|_2 = O_P(T^{-1/2}), \quad |\widehat{\phi} - \phi_0|_2 = O_P(T^{-(1-2\varphi)}),$$

which is tight and identical to the case of the known factor, but not so when $\sqrt{N} = o(T^{1-2\varphi})$.

Proof. As $\widehat{\alpha}$ and $\widehat{\gamma}$ are minimizers of $\widetilde{\mathbb{S}}_T$,

$$0 \geq \widetilde{\mathbb{S}}_T(\widehat{\alpha}, \widehat{\gamma}) - \widetilde{\mathbb{S}}_T(\alpha_0, \gamma_0) = \widetilde{\mathbb{R}}_T(\widehat{\alpha}, \widehat{\gamma}) - \widetilde{\mathbb{G}}_T(\widehat{\alpha}, \widehat{\gamma}) + \widetilde{\mathbb{G}}_T(\alpha_0, \gamma_0),$$

So $\widetilde{R}_1(\widehat{\alpha}, \widehat{\gamma}) + \widetilde{R}_2(\widehat{\gamma}) + \widetilde{\mathbb{C}}_3(\widehat{\delta}, \widehat{\gamma}) + \widetilde{R}_3(\widehat{\alpha}, \widehat{\gamma}) \leq \widetilde{\mathbb{C}}_1(\widehat{\delta}, \widehat{\gamma}) + \widetilde{\mathbb{C}}_2(\widehat{\alpha}) - \widetilde{\mathbb{C}}_4(\widehat{\alpha})$. By Lemma G.1,

$$\begin{aligned}
&\mathbb{R}(\alpha, \widehat{\phi}) + \widehat{R}_2(\widehat{\gamma}) + \widehat{\mathbb{C}}_3(\widehat{\delta}, \widehat{\gamma}) + \widehat{R}_3(\widehat{\alpha}, \widehat{\gamma}) \leq o_P(1)|\widehat{\alpha} - \alpha_0|_2^2 + O_P(\Delta_f + T^{-6})T^{-\varphi} \\
&+O_P(\Delta_f + T^{-1/2})|\widehat{\alpha} - \alpha_0|_2 + \widehat{\mathbb{C}}_1(\widehat{\delta}, \widehat{\gamma}).
\end{aligned}$$

Note that $\mathbb{R}_3(\alpha, \phi) = \widehat{R}_3(\alpha, H_T^{-1}\phi)$, $\mathbb{R}_2(\phi) = \widehat{R}_2(H_T^{-1}\phi)$, $\mathbb{C}_i(\delta, \phi) = \widehat{\mathbb{C}}_i(\delta, H_T^{-1}\phi)$, $i = 1, 3$. In addition, since $\varphi < 1/2$, by Lemma G.2, it follows that there is $C_1 > 0$,

$$\begin{aligned}
&\mathbb{R}(\alpha, \widehat{\phi}) + \mathbb{R}_2(\widehat{\phi}) + \mathbb{C}_3(\delta_0, \widehat{\phi}) \leq o_P(1)|\widehat{\alpha} - \alpha_0|_2^2 + O_P(\Delta_f + T^{-6})T^{-\varphi} \\
&+\mathbb{C}_1(\delta_0, \widehat{\phi}) + O_P(\Delta_f + T^{-1/2} + T^{-\varphi}N^{-1/2})|\widehat{\alpha} - \alpha_0|_2 + C_1 T^{-\varphi} \left|\widehat{\phi} - \phi_0\right|_2 |\widehat{\alpha} - \alpha_0|_2.
\end{aligned}$$

We now provide a lower bound on the left hand side. By Lemma G.3, $|\mathbb{R}_T(\widehat{\alpha}, \widehat{\phi}) - \mathbb{R}_T(\widehat{\alpha}, \phi_0)| = o_P(1)|\widehat{\alpha} - \alpha_0|_2^2$. Also, uniformly in α,

$$\mathbb{R}(\alpha, \phi) = \mathbb{E}[(\alpha - \alpha_0)'\mathbb{Z}_t(\phi)]^2 \geq C|\alpha - \alpha_0|_2^2.$$

In addition, $\mathbb{R}_2(\widehat{\phi}) + \mathbb{C}_3(\delta_0, \widehat{\phi}) \geq \mathbb{G}_1(\widehat{\phi}) - \mathbb{G}_2(\widehat{\phi})$. This implies

$$\begin{aligned}
&(C_0 - o_P(1))|\widehat{\alpha} - \alpha_0|_2^2 + \mathbb{G}_1(\widehat{\phi}) \leq \mathbb{G}_2(\widehat{\phi}) + \mathbb{C}_1(\delta_0, \widehat{\phi}) + O_P(\Delta_f + T^{-6})T^{-\varphi} \\
&+O_P(\Delta_f + T^{-1/2} + T^{-\varphi}N^{-1/2})|\widehat{\alpha} - \alpha_0|_2 + C_1 T^{-\varphi} \left|\widehat{\phi} - \phi_0\right|_2 |\widehat{\alpha} - \alpha_0|_2. \qquad \text{(G.13)}
\end{aligned}$$

Let C_3 be chosen to be smaller than $C_0/2$ and C_2 be chosen to be smaller than $C_4/4$ below. Due to the consistency of $\widehat{\phi}$, with probability approaching one, $|\widehat{\phi} - \phi_0|_2 \leq (C_2C_3)/(8C_1^2)$. Hence with probability approaching one, for $d = \frac{C_3}{4C_1^2}$, one term on the right hand side:

$$\begin{aligned}
&C_1 T^{-\varphi} \left|\widehat{\phi} - \phi_0\right|_2 |\widehat{\alpha} - \alpha_0|_2 \leq C_1^2 d|\widehat{\alpha} - \alpha_0|_2^2 + T^{-2\varphi} \left|\widehat{\phi} - \phi_0\right|_2^2 d^{-1} \\
\leq\ & C_3|\widehat{\alpha} - \alpha_0|_2^2/4 + C_2 T^{-2\varphi} \left|\widehat{\phi} - \phi_0\right|_2 /2.
\end{aligned}$$

Given this, the goal becomes lower bounding $\mathbb{G}_1(\widehat{\phi})$ and upper bounding $\mathbb{G}_2(\widehat{\phi}) + \mathbb{C}_1(\delta_0, \widehat{\phi})$.

Apply Lemma G.4 using the upper bound b_{NT}, and reach,

$$\mathbb{G}_2(\widehat{\phi}) + \mathbb{C}_1(\delta_0, \widehat{\phi}) \leq O_P(1) b_{NT} \leq O_P(T^{-1}) + \eta T^{-2\varphi} \left| \widehat{\phi} - \phi_0 \right|_2 .$$

with an arbitrarily small $\eta > 0$. Lemma G.5 implies $\mathbb{G}_1(\widehat{\phi}) \geq C_4 T^{-2\varphi} |\widehat{\phi} - \phi_0|_2 - \frac{C}{\sqrt{N} T^{2\varphi}}$ almost surely. Since $\eta > 0$ is arbitrarily small, (G.13) implies,

$$\begin{aligned} & C_0 |\widehat{\alpha} - \alpha_0|_2^2 / 4 + C_4 T^{-2\varphi} |\widehat{\phi} - \phi_0|_2 / 2 \\ \leq \quad & O_P(T^{-1} + \frac{C}{\sqrt{N} T^{2\varphi}}) + O_P(\Delta_f + T^{-1/2} + T^{-\varphi} N^{-1/2}) |\widehat{\alpha} - \alpha_0|_2 + O_P(\Delta_f + T^{-6}) T^{-\varphi} \end{aligned} \tag{G.14}$$

which leads to the preliminary rate: when $T^{2\varphi} \log^{\kappa} T = O(N)$ for any $\kappa > 0$,

$$\begin{aligned} |\widehat{\alpha} - \alpha_0|_2 &= O_P(T^{-1/2} + N^{-1/4} T^{-\varphi} + \Delta_f^{1/2} T^{-\varphi/2} + \Delta_f) = O_P(T^{-1/2} + N^{-1/4} T^{-\varphi}), \\ |\widehat{\phi} - \phi_0|_2 &= O_P(T^{-(1-2\varphi)} + N^{-1/2} + \Delta_f T^{\varphi} + (\Delta_f T^{\varphi})^2) = O_P(T^{-(1-2\varphi)} + N^{-1/2}), \end{aligned}$$

where we used $\Delta_f \leq O(\log^c T)(\frac{1}{N} + \frac{1}{T})$ proved in Proposition G.1. ∎

To improve the convergence rate when $N = o(T^{2-4\varphi})$, we need to obtain a sharper lower bound for $\mathbb{G}_1(\phi)$ than that of Lemma G.5 . To present the lemma below, we first introduce some notation. Let $p_{X_t|Y_t}$ denote the conditional density of X_t given Y_t, for the random vectors X_t and Y_t specified in the lemma below, assumed to exist.

Lemma G.6. *Let $u_t = g_t' \phi_0$ and Assumption 4.5 hold. Suppose $N = o(T^{2-4\varphi})$. Consider a generic deterministic vector ϕ that is linearly independent of ϕ_0 and $\sqrt{N} |\phi - \phi_0| \leq L$ for some $L > 0$. Then uniformly in ϕ,*

$$|\mathbb{G}_1(\phi) \quad \geq \quad C T^{-2\varphi} \sqrt{N} |\phi - \phi_0|_2^2 - O(\frac{1}{T^{2\varphi} N^{5/6}}).$$

Proof. Write $1_t = 1\{g_t' \phi_0 > 0\}$. First, we note that a careful calculation yields:

$$\begin{aligned} & 2 \left(1\{\breve{g}_t' \phi_0 > 0\} - 1_t\right) \left(1\{\breve{g}_t' \phi > 0\}) - 1\{\breve{g}_t' \phi_0 > 0\}\right) + \left|1\{\breve{g}_t' \phi > 0\} - 1\{\breve{g}_t' \phi_0 > 0\}\right| \\ := \quad & A_{1t}(\phi) + A_{2t}(\phi) - A_{3t}(\phi) - A_{4t}(\phi) \end{aligned}$$

where

$$\begin{aligned} A_{1t}(\phi) &= 1\left\{\breve{g}_t' \phi \leq 0 < \breve{g}_t' \phi_0\right\} 1\left\{g_t' \phi_0 > 0\right\} \\ A_{2t}(\phi) &= 1\left\{\breve{g}_t' \phi_0 \leq 0 < \breve{g}_t' \phi\right\} 1\left\{g_t' \phi_0 \leq 0\right\} \\ A_{3t}(\phi) &= 1\left\{\breve{g}_t' \phi \leq 0 < \breve{g}_t' \phi_0\right\} 1\left\{g_t' \phi_0 \leq 0\right\} \\ A_{4t}(\phi) &= 1\left\{\breve{g}_t' \phi_0 \leq 0 < \breve{g}_t' \phi\right\} 1\left\{g_t' \phi_0 > 0\right\} \end{aligned}$$

Therefore,

$$\mathbb{G}_1(\phi) \quad = \quad \mathbb{E}\left(x_t'\delta_0\right)^2\left(A_{1t}\left(\phi\right) + A_{2t}\left(\phi\right) - A_{3t}\left(\phi\right) - A_{4t}\left(\phi\right)\right).$$

The goal is to provide a sharp lower bound of the right hand side. Note that $\phi - \phi_0$ is linearly independent of ϕ_0 due to the normalization. And as elsewhere C is a generic positive constant.

Calculating A_1

Take the first term $A_{1t}\left(\phi\right)$ and note that (cf. notation $u_t = g_t'\phi_0$)

$$\begin{aligned}
A_1 &= 1\left\{0 \vee -\frac{h_t'\phi_0}{\sqrt{N}} < u_t \leq -\left(g_t + \frac{h_t}{\sqrt{N}}\right)'\left(\phi - \phi_0\right) - \frac{h_t'\phi_0}{\sqrt{N}}\right\} \\
&= 1\left\{-h_t'\phi_0 < \sqrt{N}u_t \leq -\sqrt{N}g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\}1\left\{h_t'\phi_0 \leq 0\right\} \\
&\quad + 1\left\{0 < \sqrt{N}u_t \leq -\sqrt{N}g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\}1\left\{h_t'\phi_0 > 0\right\} \\
&\quad + \left[1\left\{\sqrt{N}u_t \leq -\sqrt{N}\breve{g}_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\} - 1\left\{\sqrt{N}u_t \leq -\sqrt{N}g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\}\right] \\
&\quad\quad \times[1\left\{h_t'\phi_0 \leq 0\right\}1\{-h_t'\phi_0 < \sqrt{N}u_t\} + 1\left\{h_t'\phi_0 > 0\right\}1\{u_t > 0\}].
\end{aligned}$$

Now suppose that for any $L > 0$, the conditional density of $g_t'\phi$ given (h_t, x_t) is bounded uniformly for $\phi \in \{|\phi - \phi_0|_2 < LN^{-1/2}\}$: that is $\sup_{|\phi-\phi_0|_2<LN^{-1/2}} p_{g_t'\phi|h_t,x_t}(\cdot) < C$. Hence

$$\begin{aligned}
\mathbb{E}\left(x_t'\delta_0\right)^2 A_1 \quad &= \quad \mathbb{E}\left(x_t'\delta_0\right)^2 1\left\{-h_t'\phi_0 < \sqrt{N}u_t \leq -\sqrt{N}g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\}1\left\{h_t'\phi_0 \leq 0\right\} \\
&\quad + \mathbb{E}\left(x_t'\delta_0\right)^2 1\left\{0 < \sqrt{N}u_t \leq -\sqrt{N}g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\}1\left\{h_t'\phi_0 > 0\right\} + A_{11},
\end{aligned}$$

where

$$\begin{aligned}
A_{11} &:= \mathbb{E}\left(x_t'\delta_0\right)^2[1\left\{h_t'\phi_0 \leq 0\right\}1\{-h_t'\phi_0 < \sqrt{N}u_t\} + 1\left\{h_t'\phi_0 > 0\right\}1\{u_t > 0\}] \\
&\quad\quad \times\left[1\left\{\sqrt{N}u_t \leq -\sqrt{N}\breve{g}_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\} - 1\left\{\sqrt{N}u_t \leq -\sqrt{N}g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\}\right] \\
&\leq CT^{-2\varphi}\mathbb{E}\mathbb{P}\left\{-h_t'\left(\phi - \phi_0\right) - h_t'\phi_0 < \sqrt{N}g_t'\phi \leq -h_t'\phi_0\middle| h_t\right\} \\
&\quad + T^{-2\varphi}\mathbb{E}\mathbb{P}\left\{-h_t'\phi_0 < \sqrt{N}g_t'\phi \leq -h_t'\left(\phi - \phi_0\right) - h_t'\phi_0\middle| h_t\right\} \\
&\leq 2C\sup_{\|\phi-\phi_0\|<LN^{-1/2}} p_{g_t'\phi|h_t}(\cdot)T^{-2\varphi}\mathbb{E}\frac{|h_t'(\phi - \phi_0)|}{\sqrt{N}} \\
&\leq \frac{C}{\sqrt{N}T^{2\varphi}}|\phi - \phi_0|_2 \leq \frac{CL}{NT^{2\varphi}}, \qquad \text{given that } |\phi - \phi_0|_2 < LN^{-1/2},
\end{aligned}$$

due to Assumption 4.5 (vi) for the first inequality. On the other hand, note that the normalization condition requires the first element of $\gamma - \gamma_0 = 0$, so $g_t'\left(\phi - \phi_0\right) = f_t'(\gamma - \gamma_0) = f_{2t}'(\gamma - \gamma_0)_2$. Thus $g_t'(\phi - \phi_0)$ depends on g_t only through $f_{2t} = (H_T'f_t)_2$, where f_{2t} and

$(H_T' f_t)_2$ denote the subvectors of f_t and $H_T' f_t$, excluding their first elements, corresponding to the 1-element of ϕ.

Let $p_{u_t|\bigstar}(\cdot) := p_{f_t'\gamma_0|h_t'\phi_0, f_{2t}, x_t}(\cdot)$ denote the conditional density of $u_t = f_t'\gamma_0 = g_t'\phi_0$, given $(h_t'\phi_0, f_{2t}, x_t)$. Change variable $a = \sqrt{N}u$, we have,

$$
\begin{aligned}
&\mathbb{E}\left(x_t'\delta_0\right)^2 A_1 - A_{11} \\
&= \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{-h_t'\phi_0 < a \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 \le 0\right\} p_{u_t|\bigstar}\left(\frac{a}{\sqrt{N}}\right)da \\
&+ \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{0 < a \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 > 0\right\} p_{u_t|\bigstar}\left(\frac{a}{\sqrt{N}}\right)da \\
&= -\mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_t|\bigstar}(0) g_t'(\phi-\phi_0) 1\{g_t'(\phi-\phi_0) \le 0\} 1\left\{h_t'\phi_0 \le 0\right\} \\
&- \mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_t|\bigstar}(0)\left(g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}}\right) 1\left\{g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}} < 0\right\} 1\left\{h_t'\phi_0 > 0\right\} \\
&+ B_1,
\end{aligned} \tag{G.15}
$$

where

$$
\begin{aligned}
B_1 = {} & \frac{\mathbb{E}\left(x_t'\delta_0\right)^2}{\sqrt{N}} \int 1\left\{-h_t'\phi_0 < a \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 \le 0\right\}\left(p_{u_t|\bigstar}\left(\frac{a}{\sqrt{N}}\right) - p_{u_t|\bigstar}(0)\right)da \\
& + \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{0 < a \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 > 0\right\}\left(p_{u_t|\bigstar}\left(\frac{a}{\sqrt{N}}\right) - p_{u_t|\bigstar}(0)\right)da.
\end{aligned}
$$

We now show that for some C independent of γ, $|B_1| \le \frac{C}{NT^{2\varphi}}$. Because $p_{u_t|\bigstar}(.)$ is Lipschitz,

$$
\begin{aligned}
|B_1| &\le \frac{C}{N}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{-h_t'\phi_0 < a \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 \le 0\right\} |a| da \\
&\quad + \frac{C}{N}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{0 < a \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 > 0\right\} |a| da \\
&\le \frac{C'T^{-2\varphi}}{N}\mathbb{E}(|\sqrt{N}g_t'(\phi-\phi_0) + h_t'\phi_0| + |h_t'\phi_0|)^2 \le \frac{C'}{N}T^{-2\varphi},
\end{aligned}
$$

due to Assumption 4.5 (vi).

Calculating A_2

The calculation of A_2 is very similar to that of A_1. Write

$$
\begin{aligned}
A_2 &= 1\left\{-\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0 < \sqrt{N}u_t \le -h_t'\phi_0\right\} 1\left\{h_t'\phi_0 > 0\right\} \\
&\quad + 1\left\{-\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0 < \sqrt{N}u_t \le 0\right\} 1\left\{h_t'\phi_0 \le 0\right\} \\
&\quad + [1\left\{-\sqrt{N}\breve{g}_t'(\phi-\phi_0) - h_t'\phi_0 < \sqrt{N}u_t\right\} - 1\left\{-\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0 < \sqrt{N}u_t\right\}] \\
&\qquad \times [1\left\{h_t'\phi_0 > 0\right\} 1\left\{\sqrt{N}u_t \le -h_t'\phi_0\right\} + 1\left\{h_t'\phi_0 \le 0\right\} 1\left\{u_t \le 0\right\}].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{E}\left(x_t'\delta_0\right)^2 A_2 &= \mathbb{E}\left(x_t'\delta_0\right)^2 1\left\{-\sqrt{N}g_t'\left(\phi-\phi_0\right)-h_t'\phi_0<\sqrt{N}u_t\leq -h_t'\phi_0\right\}1\left\{h_t'\phi_0>0\right\} \\
&\quad+\mathbb{E}\left(x_t'\delta_0\right)^2 1\left\{-\sqrt{N}g_t'\left(\phi-\phi_0\right)-h_t'\phi_0<\sqrt{N}u_t\leq 0\right\}1\left\{h_t'\phi_0\leq 0\right\}+A_{21} \\
A_{21} &:= \mathbb{E}\left(x_t'\delta_0\right)^2\left[1\left\{h_t'\phi_0>0\right\}1\left\{\sqrt{N}u_t\leq -h_t'\phi_0\right\}+1\left\{h_t'\phi_0\leq 0\right\}1\left\{u_t\leq 0\right\}\right] \\
&\quad\times[1\left\{-\sqrt{N}\breve{g}_t'\left(\phi-\phi_0\right)-h_t'\phi_0<\sqrt{N}u_t\right\}-1\left\{-\sqrt{N}g_t'\left(\phi-\phi_0\right)-h_t'\phi_0<\sqrt{N}u_t\right\}] \\
&\leq \frac{CL}{NT^{2\varphi}}, \quad \text{similar to the bound of } A_{11}.
\end{aligned}
$$

So very similar to the bound of $\mathbb{E}\left(x_t'\delta_0\right)^2 A_1 - A_{11}$, we have

$$
\begin{aligned}
& \mathbb{E}\left(x_t'\delta_0\right)^2 A_2 - A_{21} \\
= \; & B_2+\mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_t|\bigstar}(0)g_t'\left(\phi-\phi_0\right)1\left\{g_t'\left(\phi-\gamma_0\right)>0\right\}1\{h_t'\phi_0>0\} \\
& +\mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_t|\bigstar}(0)\left(g_t'\left(\phi-\phi_0\right)+\frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'\left(\phi-\phi_0\right)+\frac{h_t'\phi_0}{\sqrt{N}}>0\right\}1\left\{h_t'\phi_0\leq 0\right\}
\end{aligned}
$$

with $|B_2| \leq \frac{C}{NT^{2\varphi}}$.

Calculating A_3

First we define events

$$
\begin{aligned}
E_1 &:= \{\sqrt{N}g_t'\phi_0 \leq -\sqrt{N}\breve{g}_t'\left(\phi-\phi_0\right)-h_t'\phi_0\} \\
E_2 &:= \{\sqrt{N}g_t'\left(\phi-\phi_0\right)+h_t'\phi_0>0\} \\
E_3 &:= \{\sqrt{N}\breve{g}_t'\left(\phi-\phi_0\right)+h_t'\phi_0>0\} \\
E_4 &:= \{\sqrt{N}g_t'\phi_0 \leq -\sqrt{N}g_t'\left(\phi-\phi_0\right)-h_t'\phi_0\} \\
E_5 &:= \{0<\sqrt{N}g_t'\left(\phi-\phi_0\right)+h_t'\phi_0<-h_t'(\phi-\phi_0)\} \\
E_6 &:= \{-h_t'(\phi-\phi_0)<\sqrt{N}g_t'\left(\phi-\phi_0\right)+h_t'\phi_0<0\}
\end{aligned}
$$

Careful calcuations yield:

$$
\begin{aligned}
A_3 &= 1\left\{\breve{g}_t'\phi\leq 0<\breve{g}_t'\phi_0\right\}1\left\{g_t'\phi_0\leq 0<\breve{g}_t'\phi_0\right\} \\
&= 1\left\{-h_t'\phi_0<\sqrt{N}g_t'\phi_0\leq 0\right\}1\{\sqrt{N}g_t'\left(\phi-\phi_0\right)+h_t'\phi_0<0\} \\
&\quad+1\left\{-h_t'\phi_0<\sqrt{N}g_t'\phi_0\leq -\sqrt{N}g_t'\left(\phi-\phi_0\right)-h_t'\phi_0\right\}1\{E_2\}+A_{31} \\
A_{31} &:= [1\left\{E_1\right\}+1\left\{\sqrt{N}g_t'\phi_0\leq 0\right\}]1\left\{-h_t'\phi_0<\sqrt{N}g_t'\phi_0\right\}[1\{E_3\}-1\{E_2\}] \\
&\quad+1\{E_2\}1\left\{-h_t'\phi_0<\sqrt{N}g_t'\phi_0\right\}\left[1\left\{E_1\right\}-1\left\{E_4\right\}\right].
\end{aligned}
$$

So

$$
\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2 A_3 \\
= \quad & \mathbb{E}(x_t'\delta_0)^2 1\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0 \leq -\sqrt{N} g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\{E_2\} \\
& + \mathbb{E}(x_t'\delta_0)^2 1\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0 \leq 0\right\} 1\{\sqrt{N} g_t'(\phi-\phi_0) + h_t'\phi_0 < 0\} + \mathbb{E}(x_t'\delta_0)^2 A_{31}.
\end{aligned}
$$

Note that $\sqrt{N}\breve{g}_t = \sqrt{N} g_t + h_t$, so $|1\{E_3\} - 1\{E_2\}| \leq 1\{E_5\} + 1\{E_6\}$. This gives, by Assumption 4.5 (vi) and letting $M_0 = 1$ to simplify the notation,

$$
\begin{aligned}
\mathbb{E}(x_t'\delta_0)^2 A_{31} \quad \leq \quad & T^{-2\varphi}\mathbb{E}[1\{E_5\} + 1\{E_6\}][1\{E_1\} + 1\left\{\sqrt{N} g_t'\phi_0 \leq 0\right\}]1\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0\right\} \\
& + T^{-2\varphi}\mathbb{E}1\left\{-h_t'(\phi-\phi_0) - h_t'\phi_0 < \sqrt{N} g_t'\phi \leq -h_t'\phi_0\right\} \\
& + T^{-2\varphi}\mathbb{E}1\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi < -h_t'(\phi-\phi_0) - h_t'\phi_0\right\} \\
\leq \quad & T^{-2\varphi}\mathbb{E}1\{E_5\}\mathbb{P}\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0 \leq -\sqrt{N}\breve{g}_t'(\phi-\phi_0) - h_t'\phi_0 \middle| h_t, g_t' r\right\} \\
& + T^{-2\varphi}\mathbb{E}1\{E_5\}\mathbb{P}\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0 < 0 \middle| h_t, g_t' r\right\} \\
& + T^{-2\varphi}\mathbb{E}1\{E_6\}\mathbb{P}\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0 \leq -\sqrt{N}\breve{g}_t'(\phi-\phi_0) - h_t'\phi_0 \middle| h_t, g_t' r\right\} \\
& + T^{-2\varphi}\mathbb{E}1\{E_6\}\mathbb{P}\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi_0 < 0 \middle| h_t, g_t' r\right\} \\
& + T^{-2\varphi}\mathbb{E}1\left\{-h_t'(\phi-\phi_0) - h_t'\phi_0 < \sqrt{N} g_t'\phi \leq -h_t'\phi_0 \middle| h_t\right\} \\
& + T^{-2\varphi}\mathbb{E}1\left\{-h_t'\phi_0 < \sqrt{N} g_t'\phi < -h_t'(\phi-\phi_0) - h_t'\phi_0 \middle| h_t\right\} \\
\leq^{(1)} \quad & T^{-2\varphi}\mathbb{E}1\{E_5\}C|\breve{g}_t'(\phi-\phi_0)| + T^{-2\varphi}\mathbb{E}\mathbb{P}\{E_5|h_t, x_t\}C|\frac{h_t'\phi_0}{\sqrt{N}}| \\
& + T^{-2\varphi}\mathbb{E}1\{E_6\}C|\breve{g}_t'(\phi-\phi_0)| + T^{-2\varphi}\mathbb{E}\mathbb{P}\{E_6|h_t, x_t\}C|\frac{h_t'\phi_0}{\sqrt{N}}| \\
& + \mathbb{E}C|\frac{h_t'(\phi-\phi_0)}{\sqrt{N}}| \\
\leq^{(2)} \quad & T^{-2\varphi}|\phi-\phi_0|_2 C(\mathbb{E}[|\breve{g}_t|]^q)^{1/q}(\mathbb{E}\mathbb{P}\{E_5|h_t\})^{1/p} \\
& + T^{-2\varphi}|\phi-\phi_0|_2 C(\mathbb{E}[|\breve{g}_t|]^q)^{1/q}(\mathbb{E}\mathbb{P}\{E_6|h_t\})^{1/p} \\
& + T^{-2\varphi}C\mathbb{E}|\frac{h_t' r}{\sqrt{N}}||\frac{h_t'\phi_0}{\sqrt{N}}| + T^{-2\varphi}\mathbb{E}C|\frac{h_t'(\phi-\phi_0)}{\sqrt{N}}| \\
\leq^{(3)} \quad & |\phi-\phi_0|_2 C(\mathbb{E}|\frac{h_t' r}{\sqrt{N}}|)^{1/p} T^{-2\varphi} + T^{-2\varphi}C\mathbb{E}|\frac{h_t' r h_t'\phi_0}{N}| + \mathbb{E}C|\frac{h_t'(\phi-\phi_0)}{\sqrt{N}}| \\
\leq^{(4)} \quad & O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}})
\end{aligned}
$$

where inequality (1) follows from the assumption that the conditional density $p_{u_t|\bigstar}$ and the conditional density of $g_t'\phi$ given (h_t) are bounded in a neighborhood of zero, with $r = |\phi-\phi_0|_2^{-1}(\phi-\phi_0)$; (2) (3) follow from the Holder's inequality for some $p > 1$ and $q > 0$

and $p^{-1}+q^{-1}=1$, and that the conditional density of $g_t'r$ given (h_t) is bounded. (We take $p=1.5$.); (4) follows from $|\phi-\phi_0|_2<LN^{-1/2}$.

Also,

$$\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2A_3-\mathbb{E}(x_t'\delta_0)^2A_{31} \\
= \quad & \mathbb{E}(x_t'\delta_0)^2\int 1\left\{-h_t'\phi_0<a\le 0\right\}1\{\sqrt{N}g_t'(\phi-\phi_0)+h_t'\phi_0<0\}p_{u_t|\bigstar}(\frac{a}{\sqrt{N}})d\frac{a}{\sqrt{N}} \\
& +\mathbb{E}(x_t'\delta_0)^2\int 1\left\{-h_t'\phi_0<a\le -\sqrt{N}g_t'(\phi-\phi_0)-h_t'\phi_0\right\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\} \\
& p_{u_t|\bigstar}(\frac{a}{\sqrt{N}})d\frac{a}{\sqrt{N}} \\
= \quad & \mathbb{E}(x_t'\delta_0)^2\int 1\left\{-h_t'\phi_0<a\le -\sqrt{N}g_t'(\phi-\phi_0)-h_t'\phi_0\right\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\}p_{u_t|\bigstar}(0)d\frac{a}{\sqrt{N}} \\
& +\mathbb{E}(x_t'\delta_0)^2\int 1\left\{-h_t'\phi_0<a\le 0\right\}1\{\sqrt{N}g_t'(\phi-\phi_0)+h_t'\phi_0<0\}p_{u_t|\bigstar}(0)d\frac{a}{\sqrt{N}}-B_3 \\
= \quad & -\mathbb{E}p_{u_t|\bigstar}(0)(x_t'\delta_0)^2g_t'(\phi-\phi_0)1\{\frac{h_t'\phi_0}{\sqrt{N}}>g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\} \\
& +\mathbb{E}p_{u_t|\bigstar}(0)(x_t'\delta_0)^2\frac{h_t'\phi_0}{\sqrt{N}}1\{\sqrt{N}g_t'(\phi-\phi_0)+h_t'\phi_0<0\}1\{h_t'\phi_0>0\}-B_3,
\end{aligned} \tag{G.16}$$

where

$$\begin{aligned}
& |B_3|\le \mathbb{E}(x_t'\delta_0)^2\int 1\left\{-h_t'\phi_0<a\le 0\right\}1\{\sqrt{N}g_t'(\phi-\phi_0)+h_t'\phi_0<0\}[p_{u_t|\bigstar}(\frac{a}{\sqrt{N}})-p_{u_t|\bigstar}(0)]d\frac{a}{\sqrt{N}} \\
& +C|\mathbb{E}(x_t'\delta_0)^2\frac{1}{N}\int 1\left\{-h_t'\phi_0<a\le[-\sqrt{N}g_t'(\phi-\phi_0)-h_t'\phi_0]\right\}1\{g_t'(\phi-\phi_0)>-\frac{h_t'\phi_0}{\sqrt{N}}\}|a|da \\
\le \quad & \frac{C}{N}\mathbb{E}(x_t'\delta_0)^2(|h_t'\phi_0|+|\sqrt{N}g_t'(\phi-\phi_0)|)^2\le\frac{C}{NT^{2\varphi}}.
\end{aligned}$$

Calculating A_4

Write

$$\begin{aligned}
A_4 &= 1\left\{\breve{g}_t'\phi_0\le 0<\breve{g}_t'\phi\right\}1\left\{\breve{g}_t'\phi_0\le 0<g_t'\phi_0\right\} \\
&= 1\left\{0<g_t'\phi_0\le-\frac{h_t'\phi_0}{\sqrt{N}}\right\}1\left\{-\breve{g}_t'(\phi-\phi_0)-\frac{h_t'\phi_0}{\sqrt{N}}<g_t'\phi_0\le-\frac{h_t'\phi_0}{\sqrt{N}}\right\} \\
&= 1\left\{0<g_t'\phi_0\le-\frac{h_t'\phi_0}{\sqrt{N}}\right\}1\{\breve{g}_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\} \\
&\quad 1\left\{-\breve{g}_t'(\phi-\phi_0)-\frac{h_t'\phi_0}{\sqrt{N}}<g_t'\phi_0\le-\frac{h_t'\phi_0}{\sqrt{N}}\right\}1\{\breve{g}_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}<0\}
\end{aligned}$$

The same proof as that of A_3 shows

$$\mathbb{E}(x_t'\delta_0)^2A_4$$

$$
\begin{aligned}
&= \mathbb{E}(x_t'\delta_0)^2(-h_t'\phi_0)1\{h_t'\phi_0<0\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\}p_{u_t|\bigstar(0)}\frac{1}{\sqrt{N}}\\
&\quad+\mathbb{E}(x_t'\delta_0)^2 g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)>0\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}<0\}p_{u_t|\bigstar(0)}\\
&\quad+O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}}).
\end{aligned}
$$

Combining the above results, we reach,

$$
\mathbb{E}(x_t'\delta_0)^2(A_1-A_3+A_2-A_4)=\sum_{d=1}^{8}\mathbb{E}[(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)a_d]+O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}}) \tag{G.17}
$$

where

$$
\begin{aligned}
a_1 &= -\left(g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}<0\right\}1\left\{h_t'\phi_0>0\right\}\\
a_2 &= -g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)\le 0\}1\left\{h_t'\phi_0\le 0\right\}\\
a_3 &= g_t'(\phi-\phi_0)1\{\frac{h_t'\phi_0}{\sqrt{N}}>g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\}\\
a_4 &= -\frac{h_t'\phi_0}{\sqrt{N}}1\{\sqrt{N}g_t'(\phi-\phi_0)+h_t'\phi_0<0\}1\{h_t'\phi_0>0\}\\
a_5 &= g_t'(\phi-\phi_0)1\left\{g_t'(\phi-\phi_0)>0\right\}1\{h_t'\phi_0>0\}\\
a_6 &= \left(g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\right\}1\left\{h_t'\phi_0\le 0\right\}\\
a_7 &= \frac{h_t'\phi_0}{\sqrt{N}}1\{h_t'\phi_0<0\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\}\\
a_8 &= -g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)>0\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}<0\}.
\end{aligned} \tag{G.18}
$$

We now further simplify the above terms by paying special attentions to terms involving a_2 and a_5:

$$
-\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)\le 0\}1\left\{h_t'\phi_0\le 0\right\} \tag{G.19}
$$

$$
\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)g_t'(\phi-\phi_0)1\left\{g_t'(\phi-\phi_0)>0\right\}1\{h_t'\phi_0>0\}. \tag{G.20}
$$

The key idea is that $1\left\{h_t'\phi_0\le 0\right\}$ and $1\{h_t'\phi_0>0\}$ can be exchanged up to an error $O(\frac{T^{-2\varphi}}{N})$. Roughly speaking, this is due to the fact that given (x_t,g_t), the conditional distribution of $h_t'\phi_0$ is approximately normal, and symmetric around zero. The conditional normality of $h_t'\phi_0$ follows from: for $\sigma^2_{h,x_t,g_t}:=\lim_{N\to\infty}\mathbb{E}((h_t'\phi_0)^2|x_t,g_t)$,

$$
h_t'\phi_0=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\lambda_i'\phi_0(\frac{1}{N}\Lambda'\Lambda)^{-1}|(x_t,g_t)\xrightarrow{d}\mathcal{Z}_t
$$

where $\mathcal{Z}_t$ is a Gaussian variable, whose conditional distribution given (x_t, g_t) is $\mathcal{N}(0, \sigma^2_{h,x_t,g_t})$. For a formal treatment, we show that $h_t'\phi_0$ in (G.19) and (G.20) can be replaced with $\mathcal{Z}_t$. Under the assumption of the lemma, we have

$$\sup_{x_t,g_t} |\mathbb{P}(h_t'\phi_0 \leq 0|x_t, g_t) - 1/2| = O(\frac{1}{\sqrt{N}}).$$

Then for (G.19), we have by Assumption 4.4 and 4.5

$$\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0) g_t'(\phi-\phi_0) 1\{g_t'(\phi-\phi_0)\leq 0\}[1\{h_t'\phi_0\leq 0\} - 1\{h_t'\phi_0 > 0\}] \\
= \ & \mathbb{E}p_{u_t|\bigstar}(0)(x_t'\delta_0)^2 g_t'(\phi-\phi_0) 1\{g_t'(\phi-\phi_0)\leq 0\}[1\{h_t'\phi_0\leq 0\} - 1/2] \\
& + \mathbb{E}p_{u_t|\bigstar}(0)(x_t'\delta_0)^2 g_t'(\phi-\phi_0) 1\{g_t'(\phi-\phi_0)\leq 0\}[1\{h_t'\phi_0 > 0\} - 1/2] \\
\leq \ & O_P\left(\frac{1}{\sqrt{N}}\right) \mathbb{E}\left(p_{u_t=0|\bigstar}(0)(x_t'\delta_0)^2 |g_t'(\phi-\phi_0)|\right) \\
= \ & O(\frac{T^{-2\varphi}}{N}), \quad \text{since } |\phi-\phi_0|_2 < LN^{-1/2}.
\end{aligned}$$

Hence (G.19) can be replaced with $\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0) a_2' + O(\frac{T^{-2\varphi}}{N})$, where

$$a_2' = g_t'(\phi-\phi_0) 1\{g_t'(\phi-\phi_0)\leq 0\} 1\{h_t'\phi_0 > 0\}.$$

Similarly, (G.20) can be replaced with $\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0) a_5' + O(\frac{T^{-2\varphi}}{N})$, where

$$a_5' = g_t'(\phi-\phi_0) 1\{g_t'(\phi-\phi_0) > 0\} 1\{h_t'\phi_0 < 0\}.$$

Hence with a careful calculation, up to $O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}})$ (which is uniform over ϕ), it can be shown that

$$\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2 (A_1 - A_3 + A_2 - A_4) \\
= \ & \mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)(a_1 + a_2' + a_3 + a_4 + a_5' + a_6 + a_7 + a_8). \\
= \ & -2\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)\left(g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}}\right) 1\left\{g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}} < 0\right\} 1\{h_t'\phi_0 > 0\} \\
& + 2\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)\left(g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}}\right) 1\left\{g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}} > 0\right\} 1\{h_t'\phi_0 \leq 0\}.
\end{aligned} \tag{G.21}$$

Let

$$R = -\frac{h_t'\phi_0}{\sqrt{N} g_t'(\phi-\phi_0)}.$$

Recall that $\sqrt{N}|\phi-\phi_0| \leq L$. Fix any $M_0 > 0$, we choose $\epsilon > 0$ so that when $|g_t|_2 < M_0$, then $|(1-\epsilon)\sqrt{N}g_t'(\phi-\phi_0)| \leq (1-\epsilon)LM_0$, so that $(1-\epsilon)\sqrt{N}g_t'(\phi-\phi_0)$ is inside the neighborhood of zero on which the conditional density of $h_t'\phi_0$ given (g_t, x_t) is bounded away from zero.

Thus almost surely,

$$\mathbb{P}\left\{0 < h_t'\phi_0 < -(1-\epsilon)\sqrt{N}g_t'(\phi-\phi_0)|x_t, g_t\right\} \geq c|\sqrt{N}g_t'(\phi-\phi_0)|.$$

So up to $O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}})$, by Assumption 4.5,

$$\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) \\
= \; & -2\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0) g_t'(\phi-\phi_0)(1-R)1\{0<R<1\}1\left\{h_t'\phi_0 > 0\right\} \\
& +2\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0) g_t'(\phi-\phi_0)(1-R)1\{0<R<1\}1\left\{h_t'\phi_0 \leq 0\right\} \\
\geq \; & -2\epsilon T^{-2\varphi}\mathbb{E}g_t'(\phi-\phi_0)1\left\{h_t'\phi_0 > 0\right\}1\{0<R<1-\epsilon\}1\{|g_t|_2 < M_0\} \\
& +2\epsilon T^{-2\varphi}\mathbb{E}g_t'(\phi-\phi_0)1\left\{h_t'\phi_0 \leq 0\right\}1\{0<R<1-\epsilon\}1\{|g_t|_2 < M_0\} \\
\geq \; & 2\epsilon T^{-2\varphi}\mathbb{E}1\left\{h_t'\phi_0 > 0\right\}1\{|g_t|_2 < M_0\}c\sqrt{N}|g_t'(\phi-\phi_0)|^2 \\
& +2\epsilon T^{-2\varphi}\mathbb{E}1\left\{h_t'\phi_0 \leq 0\right\}1\{|g_t|_2 < M_0\}c\sqrt{N}|g_t'(\phi-\phi_0)|^2 \\
= \; & 2c\epsilon T^{-2\varphi}\sqrt{N}\mathbb{E}|g_t'(\phi-\phi_0)|^2 1\{|g_t|_2 < M_0\} \\
\geq \; & CT^{-2\varphi}\sqrt{N}|\phi-\phi_0|_2^2,
\end{aligned}$$

where the last ineqaulity follows since the minimum eigenvalue of $\mathbb{E}\left(x_t'd_0\right)^2 g_t g_t' 1\{|g_t|_2 < M_0\}$ is bounded away from zero. It then implies

$$\mathbb{E}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) \geq C\sqrt{N}T^{-2\varphi}|\phi-\phi_0|_2^2 - O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}}), \quad p = 1.5.$$

∎

Proposition G.4. *Suppose $T = O(N)$, the first components of $\gamma_0, \widehat{\gamma}$ are one.*

$$|\widehat{\phi}-\phi_0|_2 \leq O_P\left(\frac{1}{T^{1-2\varphi}} + \frac{1}{(NT^{1-2\varphi})^{1/3}}\right).$$

Proof. Proposition G.3 shows $|\widehat{\phi}-\phi_0|_2 = O_P(T^{-(1-2\varphi)} + N^{-1/2})$. When $T^{1-2\varphi} = O(\sqrt{N})$, the above upper bound leads to

$$|\widehat{\phi}-\phi_0|_2 \leq O_P(\frac{1}{T^{1-2\varphi}}). \tag{G.22}$$

When $\sqrt{N} = O(T^{1-2\varphi})$, the above upper bound leads to $|\phi-\phi_0|_2 \leq O_P(\frac{1}{\sqrt{N}})$. We now improve this bound in the case $\sqrt{N} = O(T^{1-2\varphi})$. In this case, For an arbitrarily small $\epsilon > 0$, there is $C_e > 0$, with probability at least $1-\epsilon$, $|\phi-\phi_0|_2 \leq \frac{C_e}{\sqrt{N}}$. We now proceed the argument conditioning on this event. We use the lower bound in Lemma G.6 for $\mathbb{G}_1(\phi) = \mathbb{E}\left(x_t'\delta_0\right)^2\left(A_{1t}(\phi) + A_{2t}(\phi) - A_{3t}(\phi) - A_{4t}(\phi)\right)$.

If $\widehat{\phi}-\phi_0$ is linearly dependent of ϕ_0, there is a scalar c_T so that $\widehat{\phi}-\phi_0 = c_T\phi_0$, implying $\widehat{\phi} = (1+c_T)\phi_0$. Let $(v)_1$ denote the first component of a vector v. Then $1 = (H_T^{-1}\widehat{\phi})_1 =$

$(H_T^{-1}\phi_0)_1(1+c_T) = 1+c_T$, implying $c_T = 0$. Hence $\widehat{\phi} = \phi_0$. Hence we only need to focus on the case that $\widehat{\phi}$ is linearly independent of ϕ_0. Then Lemma G.6 yields, for $p = 1.5$

$$\mathbb{G}_1(\widehat{\phi}) \geq CT^{-2\varphi}\sqrt{N}|\widehat{\phi}-\phi_0|_2^2 - O(\frac{1}{T^{2\varphi}N^{5/6}}).$$

Write

$$m_{NT} := T^{-2\varphi}\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}.$$

Substitute to (G.13), there are $C_1, C_2, C_3 > 0$,

$$\begin{aligned}
& C|\widehat{\alpha}-\alpha_0|_2^2 + CT^{-2\varphi}\sqrt{N}|\widehat{\phi}-\phi_0|_2^2 \\
\leq\ & \mathbb{G}_2(\widehat{\phi}) + \mathbb{C}_1(\delta_0,\widehat{\phi}) + O_P(\Delta_f + T^{-6})T^{-\varphi} + O_P(\Delta_f + T^{-1/2} + T^{-\varphi}N^{-1/2})|\widehat{\alpha}-\alpha_0|_2 \\
& + C_1T^{-\varphi}\left|\widehat{\phi}-\phi_0\right|_2 |\widehat{\alpha}-\alpha_0|_2 + O(\frac{1}{T^{2\varphi}N^{5/6}}).
\end{aligned}$$

Next, replaced $\mathbb{G}_2$ and $\mathbb{C}_1$ with their upper bound based on a_{NT} given in Lemma G.4. In addition, $C_1T^{-\varphi}\left|\widehat{\phi}-\phi_0\right|_2 |\widehat{\alpha}-\alpha_0|_2 \leq C_1^2T^{-2\varphi}|\widehat{\phi}-\phi_0|_2^2N^{1/4} + |\widehat{\alpha}-\alpha_0|_2^2N^{-1/4}$. Also note that $\frac{1}{T^{2\varphi}N^{5/6}} = O(m_{NT})$ as $T = O(N)$, and $T^{-1} = O\left(m_{NT}\right)$ when $\sqrt{N} = O(T^{1-2\varphi})$.

$$\begin{aligned}
& C|\widehat{\alpha}-\alpha_0|_2^2/2 + CT^{-2\varphi}\sqrt{N}|\widehat{\phi}-\phi_0|_2^2/2 \\
\leq\ & O_P(T^{-1/2} + \Delta_f + T^{-\varphi}N^{-1/2})|\widehat{\alpha}-\alpha_0|_2 + O_P(\Delta_f + T^{-6})T^{-\varphi} + O_P\left(m_{NT}\right) \\
\leq\ & O_P(T^{-1/2} + \Delta_f)|\widehat{\alpha}-\alpha_0|_2 + O_P(m_{NT} + \Delta_f T^{-\varphi}).
\end{aligned}$$

This implies $|\widehat{\alpha}-\alpha_0|_2^2 \leq O_P(m_{NT} + \Delta_f T^{-\varphi})$ with $T^{\varphi}\log^{\kappa} T = O(N)$ for any $\kappa > 0$. Hence

$$\begin{aligned}
T^{-2\varphi}\sqrt{N}|\widehat{\phi}-\phi_0|_2^2 &\leq O_P(m_{NT} + T^{-1/2}\Delta_f^{1/2}T^{-\varphi/2} + \Delta_f\sqrt{m_{NT}} + \Delta_f^{3/2}T^{-\varphi/2} + \Delta_f T^{-\varphi}) \\
&\leq O_P(m_{NT})
\end{aligned}$$

where in the second inequality we assumed $T = O(N)$.

Hence

$$|\widehat{\phi}-\phi_0|_2^2 = O_P(T^{2\varphi}N^{-1/2}m_{NT}) = O_P\left(\frac{1}{(NT^{1-2\varphi})^{1/3}}\right)^2.$$

Combining with (G.22), we reach

$$|\widehat{\phi}-\phi_0|_2 \leq O_P\left(\frac{1}{T^{1-2\varphi}} + \frac{1}{(NT^{1-2\varphi})^{1/3}}\right).$$

∎

G.5 Consistency of Regime Classification (Proof of Theorem 4.2)

Proof of Theorem 4.2. To begin with, we consider the case of observed factors, $\widehat{f}_t = g_t$, for which we have $\phi_0 = \gamma_0$ and $\widehat{\gamma} - \gamma_0 = O_P\left(T^{-1+2\varphi}\right)$. Then, it suffices to show that

$$\sup_{|\gamma-\gamma_0|\leq CT^{-1+2\varphi}} \frac{1}{T}\sum_{t=1}^{T} \left|1\left\{g_t'\gamma > 0\right\} - 1\left\{g_t'\gamma_0 > 0\right\}\right| = O_P\left(T^{-1+2\varphi}\right),$$

for any $C < \infty$. It follows by noting that for any γ satisfying the normalization of $\gamma_1 = 1$ and for some finite c,

$$\begin{aligned} & \mathbb{E}\left|1\left\{g_t'\gamma > 0\right\} - 1\left\{g_t'\gamma_0 > 0\right\}\right| \\ = \; & \mathbb{E}\mathbb{P}\left[\left(g_{2t}'\gamma_{20} < -g_{1t} \leq g_{2t}'\gamma_2\right)|g_{1t}\right] + \mathbb{E}\mathbb{P}\left[\left(g_{2t}'\gamma_{20} \geq -g_{1t} > g_{2t}'\gamma_2\right)|g_{1t}\right] \\ \leq \; & c\mathbb{E}\left|g_{2t}'\left(\gamma_2 - \gamma_{20}\right)\right| \\ = \; & O\left(|\gamma-\gamma_0|_2\right), \end{aligned}$$

and

$$\begin{aligned} & \sup_{|\gamma-\gamma_0|_2\leq CT^{-1+2\varphi}} \left|\frac{1}{T}\sum_{t=1}^{T}\left(\left|1\left\{g_t'\gamma > 0\right\} - 1\left\{g_t'\gamma_0 > 0\right\}\right| - \mathbb{E}\left|1\left\{g_t'\gamma > 0\right\} - 1\left\{g_t'\gamma_0 > 0\right\}\right|\right)\right| \\ = \; & O_P\left(T^{-1+\varphi}\right) \end{aligned}$$

by the maximal inequality in Lemma I.1 and the subsequent remark.

Next, we move to the case of estimated factors. Recall that $\widehat{f}_t = H_T'g_t + H_T h_t/\sqrt{N}$. By the triangle inequality, for any γ

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{g_t'\phi_0 > 0\right\}\right| \leq \; & \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma > 0\right\} - 1\left\{\widetilde{f}_t'\gamma > 0\right\}\right| \\ & + \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma_0 > 0\right\} - 1\left\{\widehat{f}_t'\gamma > 0\right\}\right| \\ & + \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma_0 > 0\right\} - 1\left\{g_t'\phi_0 > 0\right\}\right|. \end{aligned} \tag{G.23}$$

Proceeding similarly as the case of the observed factors, we get

$$\frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma_0 > 0\right\} - 1\left\{\widehat{f}_t'\widehat{\gamma} > 0\right\}\right| = O_P\left(\frac{\sqrt{|\widehat{\gamma}-\gamma_0|_2}}{\sqrt{T}} + |\widehat{\gamma}-\gamma_0|_2\right)$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma_0>0\right\}-1\left\{g_t'\phi_0>0\right\}\right| = \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{g_t'\phi_0>-h_t'\phi_0/\sqrt{N}\right\}-1\left\{g_t'\phi_0>0\right\}\right| = O_P\left(\frac{1}{\sqrt{N}}\right).$$

For the remaining term in (G.23), note that

$$\begin{aligned}&\sup_{\gamma}|\frac{1}{T}\sum_{t=1}^{T}\left|1\{\widetilde{f}_t'\gamma>0\}-1\{\widehat{f}_t'\gamma>0\}\right|\\ \le\ &\sup_{\gamma}\frac{1}{T}\sum_{t=1}^{T}1\{\widehat{f}_t'\gamma<0<\widetilde{f}_t'\gamma\}+\sup_{\gamma}\frac{1}{T}\sum_{t=1}^{T}1\{\widetilde{f}_t'\gamma<0<\widehat{f}_t'\gamma\}\end{aligned}$$

and that

$$\begin{aligned}&\sup_{|\gamma|_2\le C}\frac{1}{T}\sum_{t=1}^{T}1\{\widehat{f}_t'\gamma<0<\widetilde{f}_t'\gamma\} \qquad \text{(G.24)}\\ =\ &\sup_{|\gamma|_2\le C}\frac{1}{T}\sum_{t=1}^{T}1\{-|\widehat{f}_t-\widetilde{f}_t|_2C<\widehat{f}_t'\gamma<0\}\\ \le\ &\sup_{|\gamma|_2\le C}\frac{1}{T}\sum_{t=1}^{T}1\left\{\left|\widehat{f}_t'\gamma\right|<C\Delta_f\right\}+\frac{1}{T}\sum_{t=1}^{T}1\{|\widehat{f}_t-\widetilde{f}_t|_2\ge\Delta_f\}\\ \le\ &\frac{1}{T}\sum_{t=1}^{T}1\left\{\inf_{|\gamma|_2\le C}\left|\widehat{f}_t'\gamma\right|<C\Delta_f\right\}+O_P(1)\,\mathbb{P}\{|\widehat{f}_t-\widetilde{f}_t|_2\ge\Delta_f\}\\ \le\ &O_P(1)\,\mathbb{P}\left(\inf_{|\gamma|_2\le C}|\widehat{f}_t'\gamma|<C\Delta_f\right)+O_P\left(T^{-6}\right)\\ \le\ &O_P(\Delta_f+T^{-6}),\end{aligned}$$

where the first inequality is by the fact that $1\{A\}1\{B\}\le 1\{A\}$ for any events A and B, and the remaining inequalities are by the law of iterated expectations, the rank condition in Assumption 4.1, and Proposition G.1. Recall in Proposition G.1 that notation Δ_f is introduced and $\Delta_f=O\left(T^{-1+2\varphi}\right)$ for any $\varphi>0$.

Putting together, and recalling that $\widehat{\gamma}-\gamma_0=O_P\left(\left(NT^{1-2\varphi}\right)^{-1/3}+T^{-1+2\varphi}\right)$, we conclude that

$$\sup_{|\gamma-\gamma_0|\le CT^{-1+2\varphi}}\frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma>0\right\}-1\left\{f_t'\gamma_0>0\right\}\right|=O_P\left(T^{-1+2\varphi}\right).$$

∎

Proof of Theorem 4.3 is divided into two subsections, one for the derivation of the asymptotic distribution of $\widehat{\alpha}$ and the other for the derivation of the asymptotic distribution of $\widehat{\gamma}$. The latter will contain the asymptotic independence proof as well.

G.6 Limiting distribution of $\widehat{\alpha}$ (Proof of Theorem 4.3: Part I)

Recall the notation that $\widehat{Z}_t(\gamma) = (x_t', x_t' 1\{\widehat{f}_t'\gamma > 0\})'$, $\widetilde{Z}_t(\gamma) = (x_t', x_t' 1\{\widetilde{f}_t'\gamma > 0\})'$ and $Z_t(\gamma) = (x_t', x_t' 1\{f_t'\gamma > 0\})'$. In this subsection, define $A = (\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})\widetilde{Z}_t(\widehat{\gamma})')^{-1}$. Then write

$$
\begin{aligned}
\widehat{\alpha} &= \left[\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})\widetilde{Z}_t(\widehat{\gamma})'\right]^{-1} \frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma}) y_t \\
&= \alpha_0 + (\frac{1}{T}\sum_t Z_t(\gamma_0)Z_t(\gamma_0)')^{-1}\frac{1}{T}\sum_t Z_t(\gamma_0)\varepsilon_t + \sum_{l=1}^{5} a_l,
\end{aligned}
$$

where

$$
\begin{aligned}
a_1 &= A\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})[Z_t(\gamma_0) - \widetilde{Z}_t(\gamma_0)]'\alpha_0, \\
a_2 &= A\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})[\widetilde{Z}_t(\gamma_0) - \widetilde{Z}_t(\widehat{\gamma})]'\alpha_0, \\
a_3 &= A\frac{1}{T}\sum_t [\widetilde{Z}_t(\widehat{\gamma}) - \widetilde{Z}_t(\gamma_0)]\varepsilon_t, \\
a_4 &= A\frac{1}{T}\sum_t [\widetilde{Z}_t(\gamma_0) - Z_t(\gamma_0)]\varepsilon_t, \\
a_5 &= \left[A - \left(\frac{1}{T}\sum_t Z_t(\gamma_0)Z_t(\gamma_0)'\right)^{-1}\right]\frac{1}{T}\sum_t Z_t(\gamma_0)\varepsilon_t.
\end{aligned}
$$

In view of Lemma G.1, the fact that $\mathbb{P}(|\widetilde{f}_t - \widehat{f}_t|_2 > C\Delta_f) \le O(T^{-6})$ implies $A - (\frac{1}{T}\sum_t Z_t(\gamma_0)Z_t(\gamma_0)')^{-1} = o_P(1)$, since $\widehat{\gamma} - \gamma_0 = o_P(1)$ and a ULLN applies. Hence $A = O_P(1)$ and $a_5 = o_P(T^{-1/2})$ by the MDS CLT. Furthermore, Lemma G.7 below implies $\sqrt{T}\sum_{l=1}^4 a_l = o_P(1)$. Hence

$$
\sqrt{T}(\widehat{\alpha} - \alpha_0) = (\frac{1}{T}\sum_t Z_t(\gamma_0)Z_t(\gamma_0)')^{-1}\frac{1}{\sqrt{T}}\sum_t Z_t(\gamma_0)\varepsilon_t + o_P(1).
$$

This leads to the desired strong oracle limiting distribution.

Define

$$
r_{NT} := \left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi}. \tag{G.25}
$$

Lemma G.7. *Suppose that $T = O(N)$, the conditional density of $f_t'\gamma_0$ given h_t, x_t is bounded a.s. and the density of $\inf_{\gamma\in\Gamma_T} |(g_t + h_t N^{-1/2})'\gamma|$ is bounded, where Γ_T is a r_{NT}^{-1}-neighborhood of γ_0. Then,*

(i) $\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})[Z_t(\gamma_0) - \widetilde{Z}_t(\gamma_0)]'\alpha_0 = o_P(T^{-1/2})$,
(ii) $\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})[\widetilde{Z}_t(\gamma_0) - \widetilde{Z}_t(\widehat{\gamma})]'\alpha_0 = o_P(T^{-1/2})$,
(iii) $\frac{1}{T}\sum_t[\widetilde{Z}_t(\widehat{\gamma}) - \widetilde{Z}_t(\gamma_0)]\varepsilon_t = o_P(T^{-1/2})$,
(iv) $\frac{1}{T}\sum_t[\widetilde{Z}_t(\gamma_0) - Z_t(\gamma_0)]\varepsilon_t = o_P(T^{-1/2})$.

Proof of Lemma G.7. (i) For each j,

$$
\begin{aligned}
&\left|\frac{1}{T}\sum_t \widetilde{Z}_{jt}(\widehat{\gamma})[Z_t(\gamma_0) - \widetilde{Z}_t(\gamma_0)]'\alpha_0\right| \\
&= \left|\frac{1}{T}\sum_t \widetilde{Z}_{jt}(\widehat{\gamma})x_t'\delta_0(1\{f_t'\gamma_0 > 0\} - 1\{\widetilde{f}_t'\gamma_0 > 0\})\right| \\
&\leq \frac{|\delta_0|_2}{T}\sum_t |x_t|_2^2|1\{f_t'\gamma_0 > 0\} - 1\{\widetilde{f}_t'\gamma_0 > 0\}| \\
&\leq \frac{|\delta_0|_2}{T}\sum_t |x_t|_2^2 1\{-|f_t - \widetilde{f}_t|_2|\gamma_0|_2 < f_t'\gamma_0 < 0\} + \frac{|\delta_0|_2}{T}\sum_t |x_t|_2^2 1\{0 < f_t'\gamma_0 < |f_t - \widetilde{f}_t|_2|\gamma_0|_2\}.
\end{aligned}
$$

We bound the first term on the right hand side, and the second term follows from a similar argument. In view of Lemma G.1 and the boundedness of the conditional density of $f_t'\gamma_0$,

$$
\begin{aligned}
&\frac{|\delta_0|_2}{\sqrt{T}}\sum_t |x_t|_2^2 1\{-|f_t - \widetilde{f}_t|_2|\gamma_0|_2 < f_t'\gamma_0 < 0\} \\
&\leq \frac{C}{T^{1/2+\varphi}}\sum_t |x_t|_2^2 1\{-C(\Delta_f + |\frac{h_t}{\sqrt{N}}|_2) < f_t'\gamma_0 < 0\} + \frac{C}{T^{1/2+\varphi}}\sum_t |x_t|_2^2 1\{|\widetilde{f}_t - \widehat{f}_t| > C\Delta_f\} \\
&\leq O_P(T^{1/2-\varphi})\mathbb{E}\left(|x_t|_2^2\mathbb{P}\{-C(\Delta_f + |\frac{h_t}{\sqrt{N}}|_2) < f_t'\gamma_0 < 0\,|h_t, x_t\}\right) + o_P(1) \\
&\leq O_P(T^{1/2-\varphi})\left(\Delta_f\mathbb{E}\left(|x_t|_2^2\right) + \mathbb{E}|x_t|_2^2|h_t|_2\frac{1}{\sqrt{N}}\right) + o_P(1) \\
&= o_P(1),
\end{aligned}
$$

provided that $T = O(N)$. Hence $\frac{1}{T}\sum_t \widetilde{Z}_t(\widehat{\gamma})[Z_t(\gamma_0) - \widetilde{Z}_t(\gamma_0)]\alpha_0 = o_P(T^{-1/2})$.

(ii) For each j,

$$
\begin{aligned}
&\left|\frac{1}{T}\sum_t \widetilde{Z}_{jt}(\widehat{\gamma})[\widetilde{Z}_t(\gamma_0)-\widetilde{Z}_t(\widehat{\gamma})]\alpha_0\right| \\
&\leq \frac{|\delta_0|_2}{T}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma_0<|\widehat{f}_t|_2|\gamma_0-\widehat{\gamma}|_2\} + \sup_{\gamma\in\Gamma_T}\frac{2\sqrt{2}|\delta_0|_2}{T}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma<|\widehat{f}_t-\widetilde{f}_t|_2|\gamma|_2\} \\
&+\frac{|\delta_0|_2}{T}\sum_t |x_t|^2 1\{-|\widehat{f}_t|_2|\gamma_0-\widehat{\gamma}|_2<\widehat{f}_t'\gamma_0<0\} \\
&+\sup_{\gamma\in\Gamma_T}\frac{2|\delta_0|_2}{T}\sum_t |x_t|^2 1\{-|\widehat{f}_t-\widetilde{f}_t|_2|\gamma|_2<\widehat{f}_t'\gamma<0\}.
\end{aligned}
$$

We bound the first two terms on the right hand side; the other two terms can be bounded similarly and thus details are omitted. Note that with probability at least $1-o(T^{-1})$, there is $c>0$, uniformly in t,

$$
|\widehat{f}_t|_2 \leq |H_T g_t|_2 + |H_T h_t|_2 N^{-1/2} < c(\log T)^c. \tag{G.26}
$$

Moreover, for any $\epsilon>0$, $\mathbb{P}\left\{|\widehat{\gamma}-\gamma_0|_2>\epsilon r_{NT}^{-1}\log T\right\}\to 0$. Thus

$$
\begin{aligned}
&\sqrt{T}\frac{|\delta_0|_2}{T}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma_0<|\widehat{f}_t|_2|\gamma_0-\widehat{\gamma}|_2\} \\
&= \frac{|\delta_0|_2}{\sqrt{T}}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma_0<c(\log T)^c|\gamma_0-\widehat{\gamma}|_2\} + o_P(1) \\
&= \frac{|\delta_0|_2}{\sqrt{T}}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma_0<c(\log T)^{c+1}\epsilon r_{NT}^{-1}\} + o_P(1).
\end{aligned}
$$

However, due to the boundedness of the conditional density of $\widehat{f}_t'\gamma_0$,

$$
\begin{aligned}
&\mathbb{E}\frac{|\delta_0|_2}{\sqrt{T}}\sum_t |x_t|^2 1\left\{0<\widehat{f}_t'\gamma_0<c'(\log T)^{c+1}r_{NT}^{-1}\right\} \\
&\leq T^{1/2-\varphi}\mathbb{E}\left[\mathbb{P}\left\{(0<\widehat{f}_t'\gamma_0<c(\log T)^{c+1}\epsilon r_{NT}^{-1})|x_t\right\}|x_t|^2\right] \\
&\leq C\epsilon T^{1/2-\varphi}(\log T)^{c+1}r_{NT}^{-1}\mathbb{E}|x_t|^2 \to 0 \text{ so long as } T^{1-2\varphi}(\log T)^{6c+1}=o(N^2).
\end{aligned}
$$

It remains to show $\sqrt{T}\sup_{\gamma\in\Gamma_T}\frac{2\sqrt{2}|\delta_0|_2}{T}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma<|\widehat{f}_t-\widetilde{f}_t|_2|\gamma|_2\}=o_P(1)$, which is similar to the proof of (i) due to the boundedness of γ and thus details are omitted.

Note that

$$
\sqrt{T}\sup_{\gamma\in\Gamma_T}\frac{2\sqrt{2}|\delta_0|_2}{T}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma<|\widehat{f}_t-\widetilde{f}_t|_2|\gamma|_2\}
$$

$$
\begin{aligned}
&\leq \sqrt{T}\sup_{\gamma\in\Gamma_T}\frac{2\sqrt{2}|\delta_0|_2}{T}\sum_t |x_t|^2 1\{0<\widehat{f}_t'\gamma<C\Delta_f\} + \sqrt{T}\sup_{\gamma\in\Gamma_T}\frac{2\sqrt{2}|\delta_0|_2}{T}\sum_t |x_t|^2 1\{|\widehat{f}_t-\widetilde{f}_t|_2>C\Delta_f\} \\
&\leq \sqrt{T}\frac{2\sqrt{2}|\delta_0|_2}{T}\sum_t |x_t|^2 1\{\inf_\gamma |\widehat{f}_t'\gamma|<C\Delta_f\} \leq O_P(T^{1/2-\varphi})\mathbb{P}(\inf_\gamma |\widehat{f}_t'\gamma|<C\Delta_f) \\
&= O_P(T^{1/2-\varphi}\Delta_f) = o_P(1).
\end{aligned}
$$

(iii) For each j,

$$
\begin{aligned}
&\left|\frac{1}{T}\sum_t[\widetilde{Z}_{jt}(\widehat{\gamma})-\widetilde{Z}_{jt}(\gamma_0)]\varepsilon_t\right| \\
&\leq \left|\frac{1}{T}\sum_t x_{jt}\varepsilon_t[1\{\widehat{f}_t'\widehat{\gamma}>0\}-1\{\widehat{f}_t'\gamma_0>0\}]\right| + 2\sup_{\gamma\in\Gamma_T}\left|\frac{1}{T}\sum_t x_{jt}\varepsilon_t[1\{\widehat{f}_t'\gamma>0\}-1\{\widetilde{f}_t'\gamma>0\}]\right|.
\end{aligned}
$$

Note that $\widehat{f}_t'\gamma = \breve{g}_t'\phi$ for $\breve{g}_t = g_t + h_t N^{-1/2}$ and $\phi = H^{-1}\gamma$, and $\breve{g}_t$ is ρ-mixing. Since $\widehat{\phi}$ is consistent, by Lemma I.1, the first term on the right hand side is bounded by: for any $\epsilon_1, \epsilon_2 > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(|\frac{1}{T}\sum_t x_{jt}\varepsilon_t[1\{\widehat{f}_t'\widehat{\gamma}>0\}-1\{\widehat{f}_t'\gamma_0>0\}]|_2 > T^{-1/2}\epsilon_1\right) \\
&\leq o(1) + \mathbb{P}\left(\sup_{|\phi-\phi_0|<\epsilon_1^2\sqrt{\epsilon_2}}|\frac{1}{T}\sum_t x_{jt}\varepsilon_t[1\{\breve{g}_t'\phi>0\}-1\{\breve{g}_t'\phi_0>0\}]|_2 > T^{-1/2}\epsilon_1\right) \\
&\leq o(1) + \frac{C\epsilon_1^4\epsilon_2}{\epsilon_1^4} \leq o(1) + C\epsilon_2.
\end{aligned}
$$

Because $\epsilon_1, \epsilon_2 > 0$ are arbitrary, the first term is $o(T^{-1/2})$.

As for the second term, by (G.8),

$$
\begin{aligned}
&\sup_{\gamma\in\Gamma_T}|\frac{1}{T}\sum_t x_{jt}\varepsilon_t[1\{\widehat{f}_t'\gamma>0\}-1\{\widetilde{f}_t'\gamma>0\}]| \\
&\leq \sup_{\gamma\in\Gamma_T}\frac{1}{T}\sum_t |x_{jt}\varepsilon_t|1\{\widetilde{f}_t'\gamma<0<\widehat{f}_t'\gamma\} + \sup_{\gamma\in\Gamma_T}\frac{1}{T}\sum_t |x_{jt}\varepsilon_t|1\{\widehat{f}_t'\gamma<0<\widetilde{f}_t'\gamma\} \\
&\leq O_P(\Delta_f + T^{-6}) = o_P(T^{-1/2}).
\end{aligned}
$$

(iv) By (G.8), for each j,

$$
\begin{aligned}
&\left|\frac{1}{T}\sum_t[\widetilde{Z}_{jt}(\gamma_0)-\widehat{Z}_{jt}(\gamma_0)]\varepsilon_t\right| \leq \left|\frac{1}{T}\sum_{t=1}^T \varepsilon_t x_{jt}1\{\widehat{f}_t'\gamma_0<0<\widetilde{f}_t'\gamma_0\}\right| \\
&+\left|\frac{1}{T}\sum_{t=1}^T \varepsilon_t x_j 1\{\widetilde{f}_t'\gamma_0<0<\widehat{f}_t'\gamma_0\}\right| \leq O_P(\Delta_f + T^{-6}) = o_P(T^{-1/2}),
\end{aligned}
$$

and

$$\frac{1}{T}\sum_t [\widehat{Z}_{jt}(\gamma_0) - Z_{jt}(\gamma_0)]\varepsilon_t = \frac{1}{T}\sum_{t=1}^T \varepsilon_t x_{jt} 1\{\widehat{f}_t'\gamma_0 < 0 < f_t'\gamma_0\} + \frac{1}{T}\sum_{t=1}^T \varepsilon_t x_{jt} 1\{f_t'\gamma_0 < 0 < \widehat{f}_t'\gamma_0\},$$

unless it is zero. Then, $\mathbb{E}\varepsilon_t x_{jt} 1\{\widehat{f}_t'\gamma_0 < 0 < f_t'\gamma_0\} = 0$ as ε_t is an MDS, while

$$\text{var}\left[\frac{1}{T}\sum_{t=1}^T \varepsilon_t x_{jt} 1\{\widehat{f}_t'\gamma_0 < 0 < f_t'\gamma_0\}\right] = \frac{1}{T^2}\sum_{t=1}^T \mathbb{E}x_{jt}^2 1\{\widehat{f}_t'\gamma_0 < 0 < f_t'\gamma_0\}\mathbb{E}[\varepsilon_t^2 | x_t, g_t, h_t] = o(T^{-1}).$$

Thus $\frac{1}{T}\sum_t [\widehat{Z}_t(\gamma_0) - Z_t(\gamma_0)]\varepsilon_t = o(T^{-1/2})$. ∎

G.7 Limiting distribution of $\widehat{\gamma}$ (Proof of Theorem 4.3: Part II)

Recall the defintion of r_{NT} in (G.25), which represents the convergence rate as a function of both N and T, and define

$$l_{NT} = \sqrt{r_{NT} T^{1+2\varphi}} \quad \text{and} \quad g = r_{NT}(\gamma - \gamma_0),$$

which are introduced so as to define a reparametrized process that reflects the convergence rate r_{NT}. Then, the following lemma shows that the estimator $\widehat{\gamma}$ can be represented by the following minimizer of the reparametrized version of the process:

$$\underset{g: g_1 = 0}{\operatorname{argmin}}\, l_{NT}\left[\widetilde{\mathbb{S}}_T\left(\alpha_0, \gamma_0 + \frac{g}{r_{NT}}\right) - \widetilde{\mathbb{S}}_T(\alpha_0, \gamma_0)\right].$$

Note that we fix the first element of g at 0 to impose the normalization restriction of $\gamma_1 = 0$.

The following lemma now presents the separability of the centered and scaled criterion function.

Lemma G.8. *Let $\alpha = \alpha_0 + bT^{-1/2}$, and $\gamma = \gamma_0 + g r_{NT}^{-1}$. Then, uniformly in b, g on any compact set,*

$$\begin{aligned} & l_{NT}\left[\widetilde{\mathbb{S}}_T(\alpha, \gamma) - \widetilde{\mathbb{S}}_T(\alpha_0, \gamma_0)\right] \\ = & -l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0, \gamma_0 + \frac{g}{r_{NT}}\right) + l_{NT}\mathbb{E}\left(\widehat{R}_2\left(\gamma_0 + \frac{g}{r_{NT}}\right) + \widehat{\mathbb{C}}_3\left(\gamma_0 + \frac{g}{r_{NT}}\right)\right) \\ & + l_{NT} T^{-1}\mathbb{E}[b' Z_t(\gamma_0)]^2 + l_{NT}\left[\widetilde{\mathbb{C}}_2(\alpha_0 + bT^{-1/2}) + \widetilde{\mathbb{C}}_4(\alpha_0 + bT^{-1/2})\right] \\ & + o_P(1). \end{aligned}$$

Furthermore, the two processes $l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0,\gamma_0+\frac{g}{r_{NT}}\right)$ *and* $l_{NT}\left[\widetilde{\mathbb{C}}_2(\alpha_0+bT^{-1/2})+\widetilde{\mathbb{C}}_4(\alpha_0+bT^{-1/2})\right]$ *are asymptotically independent.*

Proof. Uniformly in γ, and $\phi=H_T\gamma$, by Lemmas G.1 and G.2

$$\begin{aligned}&|\widetilde{\mathbb{C}}_1(\delta,\gamma)-\widehat{\mathbb{C}}_1(\delta_0,\gamma)|\leq|\widetilde{\mathbb{C}}_1(\delta,\gamma)-\widehat{\mathbb{C}}_1(\delta,\gamma)|+|\widehat{\mathbb{C}}_1(\delta,\gamma)-\widehat{\mathbb{C}}_1(\delta_0,\gamma)|\\ \leq\;&(T^{-\varphi}+|\alpha-\alpha_0|_2)O_P(\Delta_f+T^{-6})+\left(O_P\left(T^{-1}\right)+\eta T^{-2\varphi}\left|\phi-\phi_0\right|\right)T^{\varphi}\left|\delta-\delta_0\right|_2\end{aligned}$$

Note that $|\widehat{\gamma}-\gamma_0|_2=O_P(r_{NT}^{-1})$. Hence Lemma G.1 implies

$$\begin{aligned}l_{NT}|\widetilde{R}_2\left(\gamma\right)-\mathbb{R}_2\left(\phi\right)| &\leq O_P(\Delta_f+T^{-6})T^{-2\varphi}l_{NT}=o_P(1)\\ l_{NT}|\widetilde{R}_3| &\leq O_P(T^{-1/2}T^{-\varphi}r_{NT}^{-1})l_{NT}=o_P(1)\\ l_{NT}|\widetilde{\mathbb{C}}_1(\delta,\gamma)-\widehat{\mathbb{C}}_1(\delta_0,\gamma)| &\leq O_P(T^{-1/2})\Delta_f l_{NT}=o_P(1)\\ l_{NT}\left|\widehat{\mathbb{C}}_3\left(\delta_0,\gamma\right)-\widetilde{\mathbb{C}}_3\left(\delta,\gamma\right)\right| &\leq l_{NT}\left|\widehat{\mathbb{C}}_3\left(\delta,\gamma\right)-\widetilde{\mathbb{C}}_3\left(\delta,\gamma\right)\right|+l_{NT}\left|\widehat{\mathbb{C}}_3\left(\delta,\gamma\right)-\widehat{\mathbb{C}}_3\left(\delta_0,\gamma\right)\right|\\ &\leq l_{NT}T^{-\varphi}O_P(\Delta_f)(T^{-\varphi}+|\alpha-\alpha_0|_2)+l_{NT}T^{-\varphi}O_P(N^{-1/2})|\alpha-\alpha_0|_2\\ &\leq o_P(1).\end{aligned}$$

In addition, recall $\mathbb{G}_2:=|\widehat{R}_2(\gamma)+\widehat{\mathbb{C}}_3(\delta_0,\gamma)-(\mathbb{E}\widehat{R}_2(\gamma)+\widehat{\mathbb{C}}_3(\delta_0,\gamma))|$. By Lemma G.4, when $T^{1-2\varphi}=O(\sqrt{N})$, $l_{NT}\mathbb{G}_2\leq(O_P(\frac{1}{T})+\eta T^{-2\varphi}\left|\gamma-\gamma_0\right|_2)T^{-\varphi}l_{NT}=o_P(1)$. When $\sqrt{N}=o(T^{1-2\varphi})$, $l_{NT}\mathbb{G}_2\leq\left[T^{-2\varphi}O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right)+T^{-2\varphi}\eta r_{NT}^2\sqrt{N}\right]T^{-\varphi}l_{NT}=o_P(1)$.

Note that, $\mathbb{R}(\alpha,\phi_0)=\mathbb{E}[b'Z_t(\gamma_0)]^2$. In addition, Lemma G.1 and Lemma G.3 show uniformly in α,γ, for any $\epsilon>0$, there is $C>0$ that does not depend on ϵ,

$$\begin{aligned}&l_{NT}|\widetilde{R}_1(\alpha,\gamma)-\mathbb{R}(\alpha,\phi_0)|\leq l_{NT}|\widetilde{R}_1(\alpha,\gamma)-\mathbb{R}(\alpha,H_T^{-1}\gamma)|\\ &+l_{NT}|\mathbb{R}(\alpha,H_T^{-1}\gamma_0)-\mathbb{R}(\alpha,H_T^{-1}\gamma)|\\ \leq\;&o_P(l_{NT})|\alpha-\alpha_0|_2^2+l_{NT}C|\alpha-\alpha_0|_2^2[o_P(1)+\epsilon]^{1/2}=o_P(l_{NT})|\alpha-\alpha_0|_2^2\\ =\;&o_P(l_{NT})T^{-1}=o_P(1)\sqrt{r_{NT}T^{-1+2\varphi}}=o_P(1).\end{aligned}$$

All the above O_P,o_P are uniform in α,g. Then uniformly in α,g, for $\gamma=\gamma_0+gr_{NT}^{-1}$,

$$\begin{aligned}&l_{NT}[\widetilde{\mathbb{S}}_T\left(\alpha,\gamma\right)-\widetilde{\mathbb{S}}_T\left(\alpha_0,\gamma_0\right)]\\ =\;&l_{NT}[\widetilde{R}_1(\alpha,\gamma)+\widetilde{R}_2(\gamma)+\widetilde{R}_3(\alpha,\gamma)-\widetilde{\mathbb{C}}_1(\delta,\gamma)-\widetilde{\mathbb{C}}_2(\alpha)+\widetilde{\mathbb{C}}_3(\delta,\gamma)+\widetilde{\mathbb{C}}_4(\alpha)]\\ =\;&o_P(1)+l_{NT}[\mathbb{E}\widehat{R}_2\left(\gamma\right)+\mathbb{E}\widehat{\mathbb{C}}_3\left(\delta_0,\gamma\right)-\widehat{\mathbb{C}}_1(\delta_0,\gamma)]+l_{NT}[\mathbb{R}\left(\alpha,\phi_0\right)-\widetilde{\mathbb{C}}_2(\alpha)+\widetilde{\mathbb{C}}_4(\alpha)]\end{aligned}$$

Turning to the last claim, first note that when $l_{NT}=o\left(T\right)$, $l_{NT}T^{-1}\mathbb{E}[b'Z_t(\gamma_0)]^2=o_P\left(1\right)$ and $l_{NT}\left[\widetilde{\mathbb{C}}_2(\alpha_0+bT^{-1/2})+\widetilde{\mathbb{C}}_4(\alpha_0+bT^{-1/2})\right]=o_P\left(1\right)$ due to the proof in Section G.6. When $l_{NT}=T$, we need to show that $l_{NT}\left[\widetilde{\mathbb{C}}_2(\alpha_0+bT^{-1/2})+\widetilde{\mathbb{C}}_4(\alpha_0+bT^{-1/2})\right]$ is asymptotically uncorrelated to $l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0,\gamma_0+\frac{g}{r_{NT}}\right)$. This follows from Lemma G.9 in the ensueing section. ∎

G.7.1 Empirical Process Part

We concern the weak convergence of the empirical process given by

$$\begin{aligned} l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0,\gamma_0+\frac{g}{r_{NT}}\right) &= l_{NT}\frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\delta_0\left(\widehat{1}_t\left(\gamma_0+\frac{g}{r_{NT}}\right)-\widehat{1}_t\left(\gamma_0\right)\right) \\ &= 2\breve{\mathbb{C}}_{11}\left(H_T g\right)-2\breve{\mathbb{C}}_{12}\left(H_T g\right), \end{aligned}$$

where $\breve{u}_t=\breve{g}_t'\phi_0$ and

$$\begin{aligned} \breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right) &= \frac{\sqrt{r_{NT}}}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t x_t' d_0 1\left\{-\breve{g}_t'\frac{\mathfrak{g}}{r_{NT}}<\breve{u}_t\le 0\right\}, \\ \breve{\mathbb{C}}_{12}\left(\mathfrak{g}\right) &= \frac{\sqrt{r_{NT}}}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t x_t' d_0 1\left\{0<\breve{u}_t\le-\breve{g}_t'\frac{\mathfrak{g}}{r_{NT}}\right\}, \end{aligned}$$

where $\mathfrak{g}$ belongs to a compact set $\mathfrak{G}$. This is because $l_{NT}T^{-1-\varphi}=\sqrt{r_{NT}/T}$, $\breve{g}_t=g_t+h_t/\sqrt{N}=H_T^{-1\prime}\widehat{f}_t$, and $\widehat{f}_t'\mathfrak{g}=\breve{g}_t'H_T\mathfrak{g}$.

We introduce this transformation to remove the randomness in H_T from the definition of the processes $\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)$ and $\breve{\mathbb{C}}_{12}\left(\mathfrak{g}\right)$ and make use of the stationarity of $\breve{g}_t$. Furthermore, in view of the extended CMT in Lemma I.4 $\breve{\mathbb{C}}_{11}\left(H_T g\right)$ and $\breve{\mathbb{C}}_{11}\left(Hg\right)$ have the same weak limit if $H_T\xrightarrow{p}H$ and H is a finite constant. Thus, it is sufficient to derive the weak convergence of $\left(\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right),\breve{\mathbb{C}}_{12}\left(\mathfrak{g}\right)\right)$ to some process, say, $\left(\mathbb{C}_{11}\left(\mathfrak{g}\right),\mathbb{C}_{12}\left(\mathfrak{g}\right)\right)$. Since $\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)$ is of the same type as $\breve{\mathbb{C}}_{12}\left(\mathfrak{g}\right)$ and there is no correlation between the two as ε_t is an mds and the two indicators are orthogonal to each other, we focus on the stochastic equicontinuity and fidi of $\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)$.

The stochastic equicontinuity of $\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)$, however, is a direct consequence of Lemma I.1 since $\breve{u}_t$ and $\breve{g}_t$ are stationary triangular arrays and thus for any finite $\mathfrak{g}$ and $\gamma=\frac{\mathfrak{g}}{r_{NT}}$ and for any $c,\epsilon>0$

$$\begin{aligned} &\mathbb{P}\left\{\sup_{|\mathfrak{h}-\mathfrak{g}|<\epsilon}\left|\breve{\mathbb{C}}_{11}\left(\mathfrak{h}\right)-\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)\right|>c\right\} \\ = \;&\mathbb{P}\left\{\sup_{|\vec{\gamma}-\gamma|<\epsilon/r_{NT}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t x_t' d_0\left(1\left\{-\breve{g}_t'\gamma<\breve{u}_t\le 0\right\}-1\left\{-\breve{g}_t'\vec{\gamma}<\breve{u}_t\le 0\right\}\right)>\frac{c}{\sqrt{r_{NT}}}\right\} \\ \le \;& C\frac{\epsilon^2}{c^4}, \end{aligned}$$

which can be made arbitrarily small by choosing ϵ small.

Turning to the fidi of $\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)$, we first check $\breve{\mathbb{C}}_{11}\left(\mathfrak{g}\right)$ satisfies the conditions to apply the

mds CLT (e.g. Hall and Heyde 1980). Specifically, let $v_t = \sqrt{r_{NT}}\varepsilon_t x_t' d_0 1\left\{-\breve{g}_t' \frac{\mathfrak{g}}{r_{NT}} < \breve{u}_t \le 0\right\}$, which is an mds as ε_t is an mds, and verify that $\max_t |v_t| = o_P\left(\sqrt{T}\right)$ and that $\frac{1}{T}\sum_{t=1}^T v_t^2$ has a proper non-degenerate probability limit. However, $T^{-2}\mathbb{E}\max_t v_t^4 \le T^{-1}\mathbb{E}v_t^4$ by the stationarity and by $\max_t |a_t| \le \sum_{t=1}^T |a_t|$ and
$T^{-1}\mathbb{E}v_t^4 = T^{-1}r_{NT}^2\mathbb{E}\left(\varepsilon_t x_t' d_0\right)^4 1\left\{-\breve{g}_t' \frac{\mathfrak{g}}{r_{NT}} < \breve{u}_t \le 0\right\} \le CT^{-1}r_{NT} = o(1)$. Furthermore, $\frac{1}{T}\sum_{t=1}^T\left(v_t^2 - Ev_t^2\right) = o_P(1)$ due to Lemma I.1. Thus, it remains to show that the limit of $\mathbb{E}v_t^2$ does not degenerate, which is shown in the following.

To that end, we first derive the following limit

$$\begin{aligned} L(\mathfrak{s},\mathfrak{g}) &= \lim_{N,T\to\infty} \mathbb{E}\left(\breve{\mathbb{C}}_{11}(\mathfrak{s}) - \breve{\mathbb{C}}_{12}(\mathfrak{s}) - \breve{\mathbb{C}}_{11}(\mathfrak{g}) + \breve{\mathbb{C}}_{12}(\mathfrak{g})\right)^2 \\ &= \lim_{N,T\to\infty} r_{NT}\mathbb{E}\eta_t^2\left|1\left\{\breve{g}_t'\left(\phi_0 + \frac{\mathfrak{s}}{r_{NT}}\right) > 0\right\} - 1\left\{\breve{g}_t'\left(\phi_0 + \frac{\mathfrak{g}}{r_{NT}}\right) > 0\right\}\right| \end{aligned}$$

for $\mathfrak{s} \neq \mathfrak{g}$ and $\eta_t = \varepsilon_t x_t' d_0$.

Note that each element $\mathfrak{g} \in \mathfrak{G}$ is linearly independent of $\phi_0 = H\gamma_0$, since $g_1 = 0$ while $\gamma_{01} = 1$. Otherwise, there is $c \neq 0$ such that $\mathfrak{g} = \mathfrak{c}\phi_0$. Then, $\mathfrak{g} = Hg = cH\gamma_0$, which in turn implies that $g = c\gamma_0$. This is a contradiction as $g_1 = 0$ while $\gamma_{01} = 1$. This allows us to apply Lemma G.9 below to conclude that

$$\begin{aligned} & r_{NT}\mathbb{E}\eta_t^2 1\left\{\breve{u}_t + \breve{g}_t'\frac{\mathfrak{s}}{r_{NT}} > 0 \ge \breve{u}_t + \breve{g}_t'\frac{\mathfrak{g}}{r_{NT}}\right\} \\ \to\ & \mathbb{E}\left[\eta_t^2\left(-g_t'\mathfrak{g} + g_t'\mathfrak{s}\right)1\left(g_t'\mathfrak{g} < g_t'\mathfrak{s}\right)|u_t = 0\right]p_u(0), \end{aligned}$$

and that

$$\begin{aligned} & r_{NT}\mathbb{E}\eta_t^2 1\left\{\breve{u}_t + \breve{g}_t'\frac{\mathfrak{s}}{r_{NT}} \le 0 < \breve{u}_t + \breve{g}_t'\frac{\mathfrak{g}}{r_{NT}}\right\} \\ \to\ & \mathbb{E}\left[\eta_t^2\left(g_t'\mathfrak{g} - g_t'\mathfrak{s}\right)1\left(g_t'\mathfrak{g} > g_t'\mathfrak{s}\right)|u_t = 0\right]p_u(0). \end{aligned}$$

Thus, we conclude that

$$L(\mathfrak{s},\mathfrak{g}) = \mathbb{E}_0\left[\eta_t^2\left|g_t'(\mathfrak{g} - \mathfrak{s})\right|\big|\, u_t = 0\right]p_u(0).$$

Putting these together, we conclude

$$l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0, \gamma_0 + \frac{g}{r_{NT}}\right) \Rightarrow 2W(g),$$

where $W(g)$ is a centered Gaussian process with the covariance kernel

$$\mathbb{E}W(g)W(s) = \frac{1}{2}\left(L(Hs,0) + L(Hg,0) - L(Hs,Hg)\right),$$

recalling that $\mathbb{E}XY = \frac{1}{2}\left(\mathbb{E}X^2 + \mathbb{E}Y^2 - \mathbb{E}(X-Y)^2\right)$ and $\breve{\mathbb{C}}_{11}(0) = 0$.

Lemma G.9. *Assume Assumption 4.5. Then,*

$$r_{NT}\mathbb{E}\eta_t^2 1\left\{\breve{u}_t + \breve{g}_t'\frac{s}{r_{NT}} > 0 \geq \breve{u}_t + \breve{g}_t'\frac{g}{r_{NT}}\right\} \to \mathbb{E}\left[\eta_t^2\left(g_t's - g_t'g\right)1\left(g_t'g < g_t's\right)|u_t = 0\right]p_{u_t}(0),$$

as $N, T \to \infty$.

Proof of Lemma G.9. First, we write a conditional density of $\breve{u}_t$ given a random variable Y by $p(u|Y)$ for more clarity. Note that

$$\begin{aligned}
& r_{NT}\mathbb{E}\eta_t^2 1\left\{\breve{u}_t + \breve{g}_t'\frac{s}{r_{NT}} > 0 \geq \breve{u}_t + \breve{g}_t'\frac{w}{r_{NT}}\right\} \\
= \; & r_{NT}\mathbb{E}\eta_t^2 1\left\{-\frac{\breve{g}_t's}{r_{NT}} < \breve{u}_t \leq -\frac{\breve{g}_t'w}{r_{NT}}\right\} \\
= \; & \mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\mathbb{E}\left(\eta_t^2|\frac{z}{r_{NT}}, \breve{g}_t's, \breve{g}_t'w\right)p\left(\frac{z}{r_{NT}}|\breve{g}_t's, \breve{g}_t'w\right)dz 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\right] \\
= \; & \mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\mathbb{E}\left(\eta_t^2|0, \breve{g}_t's, \breve{g}_t'w\right)p\left(0|\breve{g}_t's, \breve{g}_t'w\right)dz 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\right] \\
& +\mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\left(\mathbb{E}\left(\eta_t^2|\frac{z}{r_{NT}}, \breve{g}_t's, \breve{g}_t'w\right) - \mathbb{E}\left(\eta_t^2|0, \breve{g}_t's, \breve{g}_t'w\right)\right)p\left(0|\breve{g}_t's, \breve{g}_t'w\right)dz 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\right] \\
& +\mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\left(p\left(\frac{z}{r_{NT}}|\breve{g}_t's, \breve{g}_t'w\right) - p\left(0|\breve{g}_t's, \breve{g}_t'w\right)\right)\mathbb{E}\left(\eta_t^2|0, \breve{g}_t's, \breve{g}_t'w\right)dz 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\right] \\
& +\mathbb{E}\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\left(\mathbb{E}\left(\eta_t^2|\frac{z}{r_{NT}}, \breve{g}_t's, \breve{g}_t'w\right) - \mathbb{E}\left(\eta_t^2|0, \breve{g}_t's, \breve{g}_t'w\right)\right)\left(p\left(\frac{z}{r_{NT}}|\breve{g}_t's, \breve{g}_t'w\right) - p\left(0|\breve{g}_t's, \breve{g}_t'w\right)\right) \\
& dz 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}
\end{aligned}$$

by a change-of-variables formula $z = r_{NT}u$. First,

$$\begin{aligned}
& \mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\mathbb{E}\left(\eta_t^2|0, \breve{g}_t's, \breve{g}_t'w\right)p\left(0|\breve{g}_t's, \breve{g}_t'w\right)dz 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\right] \\
= \; & \mathbb{E}\left(1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\left(\breve{g}_t's - \breve{g}_t'w\right)\mathbb{E}\left(\eta_t^2|0, \breve{g}_t's, \breve{g}_t'w\right)p\left(0|\breve{g}_t's, \breve{g}_t'w\right)\right) \\
= \; & \mathbb{E}\left(\eta_t^2 1\left\{\breve{g}_t's > \breve{g}_t'w\right\}\left(\breve{g}_t's - \breve{g}_t'w\right)|\breve{u}_t = 0\right)p_{\breve{u}_t}(0) \\
\to \; & \mathbb{E}\left(\eta_t^2 1\left\{g_t's > g_t'w\right\}\left(g_t's - g_t'w\right)|u_t = 0\right)p_{u_t}(0),
\end{aligned}$$

where the convergence holds by the following reasons. Since $(\eta_t, \breve{g}_t')' \xrightarrow{p} (\eta_t, g_t')'$ as $N \to \infty$, we have $\eta_t^2 1\{\breve{g}_t's > \breve{g}_t'w\}(\breve{g}_t's - \breve{g}_t'w) \xrightarrow{p} \eta_t^2 1\{g_t's > g_t'w\}(g_t's - g_t'w)$ and $\breve{u}_t \xrightarrow{p} u_t$ by the continuous mapping theorem, which imples by the Lipschitz continuity of the densities (Assumption 4.5 (vii)) the convergence of $p_{\breve{u}_t}(0)$ and the conditional densities. This in turn implies the convergence of $\mathbb{E}\left(\eta_t^2 1\{\breve{g}_t's > \breve{g}_t'w\}(\breve{g}_t's - \breve{g}_t'w)|\breve{u}_t = 0\right)$ due to the uniform integrability, which is implied by the boundedness of $\mathbb{E}\left(\eta_t^4 \left|\breve{g}_t\right|_2^2 |\breve{u}_t\right)$.

Then, we show the other terms are negligible. We elaborate the first of these since the reasonings are similar.

$$
\begin{aligned}
& \mathbb{E}\left[\int_{-\breve{g}_t' s}^{-\breve{g}_t' w}\left(\mathbb{E}\left(\eta_t^2 | \frac{z}{r_{NT}}, \breve{g}_t' s, \breve{g}_t' w\right)-\mathbb{E}\left(\eta_t^2 | 0, \breve{g}_t' s, \breve{g}_t' w\right)\right) \frac{z}{r_{NT}} p\left(0 | \breve{g}_t' s, \breve{g}_t' w\right) d z 1\left\{\breve{g}_t' s>\breve{g}_t' w\right\}\right] \\
\leq\; & C \mathbb{E}\left[\int_{-\breve{g}_t' s}^{-\breve{g}_t' w} \frac{z}{r_{NT}} d z p\left(0 | \breve{g}_t' s, \breve{g}_t' w\right) 1\left\{\breve{g}_t' s>\breve{g}_t' w\right\}\right] \\
=\; & C' \mathbb{E}\left(\left(\breve{g}_t' w\right)^2-\left(\breve{g}_t' s\right)^2\right) \frac{1}{2 r_{NT}}=o(1) .
\end{aligned}
$$

∎

G.7.2 Bias

We show that, as $N, T \rightarrow \infty$,

$$
l_{NT}(\mathbb{E} \widehat{R}_2(g)+\widehat{\mathbb{C}}_3(g)) \rightarrow A(\omega, g),
$$

where

$$
A(\omega, g):=M_\omega \mathbb{E}\left(\left(x_t' d_0\right)^2\left[\left|g_t' H g+\zeta_\omega^{-1} \mathcal{Z}_t\right|-\left|\zeta_\omega^{-1} \mathcal{Z}_t\right|\right] \middle| u_t=0\right) p_{u_t}(0).
$$

and that $A(\omega, g) \rightarrow+\infty$ as $|g| \rightarrow+\infty$ for any ω.

Proof. For $\gamma=H^{-1} \phi$, and $g=r_{NT}\left[\gamma-\gamma_0\right]$, we have $\phi-\phi_0=H\left(\gamma-\gamma_0\right)=r_{NT}^{-1} H g$, with $g_1=0$ due to the normalization. Suppose $g \neq 0$. Let

$$
r_g=\left|\phi-\phi_0\right|_2^{-1}\left(\phi-\phi_0\right)=|H g|_2^{-1} H g .
$$

We only need to focus on the case that r_g is linearly independent of ϕ_0. Let

$$
\zeta_{NT}=\sqrt{N} r_{NT}^{-1} .
$$

By the proof of Lemma G.6,

$$
\begin{aligned}
& l_{NT} \mathbb{E}\left(\widehat{\mathbb{C}}_3\left(\delta_0, \gamma\right)+\widehat{R}_2(\gamma)\right) \\
=\; & l_{NT} \mathbb{E}\left(x_t' \delta_0\right)^2\left(A_{1t}(\phi)+A_{2t}(\phi)-A_{3t}(\phi)-A_{4t}(\phi)\right)
\end{aligned}
$$

Step I: obtaining the results for the case of $\omega \in(0, \infty]$.

In this case, $\zeta_{NT} \to \zeta_\omega \in (0, \infty]$. We now work with (G.18). Note that for $p = 1.5$,

$$\frac{l_{NT}}{T^{2\varphi} N^{0.5+1/(2p)}} = o(1),$$

and

$$M_{NT} := \frac{1}{\sqrt{N}} l_{NT} T^{-2\varphi} \zeta_{NT} \to M_\omega := \max\{1, \omega^{-1/3}\} \in (0, \infty).$$

We shall use the following equality, which can be verified:

$$\begin{aligned} |a+b| - |b| &= \Xi(a,b), \quad \text{where} \\ \Xi(a,b) &:= -a1\{a \le 0\}1\{b \le 0\} - (a+b)1\{a+b<0\}1\{b>0\} \\ &\quad -b1\{a+b<0\}1\{b>0\} + a1\{a+b>0\}1\{a<0\} \\ &\quad +a1\{a>0\}1\{b>0\} + (a+b)1\{a+b>0\}1\{b \le 0\} \\ &\quad +b1\{b<0\}1\{a+b>0\} - a1\{a>0\}1\{a+b<0\}. \end{aligned} \tag{G.27}$$

Let $g_t'(\phi - \phi_0) = a$, $\frac{h_t'\phi_0}{\sqrt{N}} = b$, Note that (G.18) can be written exactly as the right hand side of the above equality, up to $\mathbb{E}_{|u_t=0}(x_t'\delta_0)^2 p_{u_t}(0)$. Hence (G.18) and the above equality imply, for $\phi - \phi_0 = r_{NT}^{-1} H_T g$,

$$\begin{aligned} & l_{NT}\mathbb{E}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) \\ =^{(1)} \;& l_{NT}\mathbb{E}_{|u_t=0}(x_t'\delta_0)^2 p_{u_t}(0)\Xi(a,b) + o(1) \\ =^{(2)} \;& l_{NT}\mathbb{E}_{|u_t=0}(x_t'\delta_0)^2 p_{u_t}(0)\left[\left|g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}}\right| - \left|\frac{h_t'\phi_0}{\sqrt{N}}\right|\right] + o(1) \\ = \;& \breve{\mathbb{C}}_{NT}(H_T g) + o(1), \quad \text{where} \\ \breve{\mathbb{C}}_{NT}(\mathfrak{g}) := \;& M_{NT}\mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\left(\left|g_t'\mathfrak{g} + \zeta_\omega^{-1}h_t'\phi_0\right| - \left|\zeta_\omega^{-1}h_t'\phi_0\right|\right) \end{aligned}$$

In the above, (1) is rewriting (G.18) using the notation of $\Xi(a,b)$ for $g_t'(\phi - \phi_0) = a$ and $\frac{h_t'\phi_0}{\sqrt{N}} = b$; (2) uses the equality $|a+b| - |b| = \Xi(a,b)$.

Step I.1: pointwise convergence of $\breve{\mathbb{C}}_{NT}(\mathfrak{g})$

We now derive the pointwise limit of $\breve{\mathbb{C}}_{NT}(\mathfrak{g})$. Define

$$\widetilde{F}_{g_t}(z) = \left|g_t'\mathfrak{g} + \zeta_\omega^{-1}z\right| - \left|\zeta_\omega^{-1}z\right|.$$

Then $\breve{\mathbb{C}}_{NT}(\mathfrak{g}) = M_{NT}\mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t, g_t]$. Now we use the following portmanteau lemma: $X_n \xrightarrow{d} X$ if and only if $\mathbb{E}\widetilde{F}(X_n) \to \mathbb{E}\widetilde{F}(X)$ for all bounded continuous functions $\widetilde{F}$. Note that $h_t'\phi_0|x_t, g_t \xrightarrow{d} \mathcal{Z}_t$. Now for each fixed (x_t, g_t),

$$|\widetilde{F}_{g_t}(z)| \le |g_t'\mathfrak{g}|;$$

the right hand side is independent of z, and $\widetilde{F}_{g_t}(z)$ is continuous in z. So we can apply the portmanteau lemma to conclude that $\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t, g_t] \to \mathbb{E}[\widetilde{F}_{g_t}(\mathcal{Z}_t)|x_t, g_t]$ for each fixed x_t, g_t. This further implies, $P_N(x_t, g_t) \to P(x_t, g_t)$ for each fixed (x_t, g_t), with

$$\begin{aligned} P_N(x_t, g_t) &:= (x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t, g_t], \\ P(x_t, g_t) &:= (x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(\mathcal{Z}_t)|x_t, g_t]. \end{aligned}$$

In addition, note that for each fixed x_t, g_t, $|\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t, g_t]| \leq |g_t'\mathfrak{g}|$. For all N, $|P_N(x_t, g_t)| \leq (x_t'd_0)^2 p_{u_t}(0)|g_t'\mathfrak{g}|$; the right hand side does not depend on N, and has a bounded expectation: $\mathbb{E}(x_t'd_0)^2 p_{u_t}(0)|g_t'\mathfrak{g}| < \infty$. Hence by the dominated convergence theorem, the pointwise convergence of $P_N(x_t, g_t) \to P(x_t, g_t)$ implies $\mathbb{E}_{|u_t=0}P_N(x_t, g_t) \to \mathbb{E}_{|u_t=0}P(x_t, g_t)$, which means

$$\mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t, g_t] \to \mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(\mathcal{Z}_t)|x_t, g_t].$$

Also, $M_{NT} \to M_\omega \in (0, \infty)$. Thus

$$\begin{aligned} \breve{\mathbb{C}}_{NT}(\mathfrak{g}) &= M_{NT}\mathbb{E}_{|u_t=0}\{(x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t, g_t]\} \\ &\to M_\omega \mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(\mathcal{Z}_t)|x_t, g_t] \\ &= M_\omega \mathbb{E}\left((x_t'd_0)^2[|g_t'\mathfrak{g} + \zeta_\omega^{-1}\mathcal{Z}_t| - |\zeta_\omega^{-1}\mathcal{Z}_t|]\middle| u_t = 0\right) p_{u_t}(0) \\ &:= \breve{A}(\mathfrak{g}). \end{aligned}$$

Hence we have proved for some $C > 0$ and any $|\mathfrak{g}|_2 < C$,

$$\begin{aligned} l_{NT}\mathbb{E}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) &= \breve{\mathbb{C}}_{NT}(H_T g) + o(1), \\ \breve{\mathbb{C}}_{NT}(\mathfrak{g}) &\to \breve{A}(\mathfrak{g}). \end{aligned}$$

Step I.2: $\breve{\mathbb{C}}_{NT}(H_T g) \xrightarrow{P} A(\omega, g)$

We apply the extended continuous mapping theorem (CMT) for drifting functions (cf. Lemma I.4). To do so, first note that $H_T \to^P H$ for some $K \times K$ invertible nonrandom matrix H (e.g., Bai (2003)). To applied the extended CMT, we need to show, for any converging sequence $\mathfrak{g}_T \to \mathfrak{g}$ in a compact space, we have

$$\breve{\mathbb{C}}_{NT}(\mathfrak{g}_T) \to \breve{A}(\mathfrak{g}). \tag{G.28}$$

Once this is achieved, then because $H_T g \xrightarrow{P} Hg$, by Theorem 1.11.1 of van der Vaart and Wellner (1996), we have $\breve{\mathbb{C}}_{NT}(H_T g) \xrightarrow{P} \breve{A}(Hg) = A(\omega, g)$.

To prove (G.28), note that $|\breve{\mathbb{C}}_{NT}(\mathfrak{g}_T) - \breve{A}(\mathfrak{g})| \leq |\breve{\mathbb{C}}_{NT}(\mathfrak{g}_T) - \breve{\mathbb{C}}_{NT}(\mathfrak{g})| + |\breve{\mathbb{C}}_{NT}(\mathfrak{g}) - \breve{A}(\mathfrak{g})|$.

The second term on the right hand side is $o(1)$ due to the pointwise convergence. It remains to prove the first term on the right is also $o(1)$. By definition,

$$
\begin{aligned}
&|\breve{\mathbb{C}}_{NT}(\mathfrak{g}_T) - \breve{\mathbb{C}}_{NT}(\mathfrak{g})| \leq M_{NT}\mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\mathbb{E}[\left|g_t'(\mathfrak{g}_T - \mathfrak{g})\right| |x_t, g_t] \\
\leq\ & O(1)\mathbb{E}_{|u_t=0}(x_t'd_0)^2|g_t|_2\,|\mathfrak{g}_T - \mathfrak{g}| \leq O(1)\,|\mathfrak{g}_T - \mathfrak{g}| = o(1).
\end{aligned}
$$

Hence by the triangular inequality, (G.28) holds. It then immediately follows that $l_{NT}\mathbb{E}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) \xrightarrow{P} A(\omega, g)$. In particular, when $\omega = \infty$, $\zeta_\omega^{-1} = 0$ and $M_\omega = 1$, so $A(\omega, g) = A(\infty, g)$.

Step II: obtaining the results for the case of $\omega = 0$

In this case, we have that $\zeta_{NT} \to 0$, and

$$
\widetilde{M}_{NT} := \frac{l_{NT}\zeta_{NT}^2}{\sqrt{N}}T^{-2\varphi} \to 1.
$$

We now work with the last equality of (G.21), up to $\frac{l_{NT}}{T^{2\varphi}N^{0.5+1/(2p)}} = o(1)$,

$$
l_{NT}\mathbb{E}_{|u_t=0}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) := \breve{\mathbb{C}}_{NT,2}(H_T g) + o(1)
$$

where

$$
\begin{aligned}
\breve{\mathbb{C}}_{NT,2}(\mathfrak{g}) \quad := \quad & -\frac{l_{NT}T^{-2\varphi}\zeta_{NT}}{\sqrt{N}}2\mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\left(g_t'\mathfrak{g} + \zeta_{NT}^{-1}h_t'\phi_0\right)1\left\{g_t'\mathfrak{g} + \zeta_{NT}^{-1}h_t'\phi_0 < 0\right\}1\left\{h_t'\phi_0 > 0\right\} \\
& + \frac{l_{NT}T^{-2\varphi}\zeta_{NT}}{\sqrt{N}}2\mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\left(g_t'\mathfrak{g} + \zeta_{NT}^{-1}h_t'\phi_0\right)1\left\{g_t'\mathfrak{g} + \zeta_{NT}^{-1}h_t'\phi_0 > 0\right\}1\left\{h_t'\phi_0 \leq 0\right\}.
\end{aligned}
$$

Step II.1: pointwise convergence of $\breve{\mathbb{C}}_{NT,2}(\mathfrak{g})$

We now derive the limit of $\breve{\mathbb{C}}_{NT,2}(\mathfrak{g})$. Change variable $y = h_t'\phi_0\zeta_{NT}^{-1}$, $\breve{\mathbb{C}}_{NT,2}(\mathfrak{g})$ equals

$$
-\widetilde{M}_{NT}2p_{u_t}(0)\mathbb{E}[(x_t'd_0)^2F_{NT,1}(g_t, x_t)|u_t = 0] + \widetilde{M}_{NT}2p_{u_t}(0)\mathbb{E}[(x_t'd_0)^2F_{NT,2}(g_t, x_t)|u_t = 0],
$$

where

$$
\begin{aligned}
F_{NT,1}(g_t, x_t) &:= \int \left(g_t'\mathfrak{g} + y\right)1\left\{g_t'\mathfrak{g} + y < 0\right\}1\left\{y > 0\right\}p_{h_t'\phi_0|g_t,x_t,u_t=0}(\zeta_{NT}y)dy \\
F_{NT,2}(g_t, x_t) &:= \int \left(g_t'\mathfrak{g} + y\right)1\left\{g_t'\mathfrak{g} + y > 0\right\}1\left\{y \leq 0\right\}p_{h_t'\phi_0|g_t,x_t,u_t=0}(\zeta_{NT}y)dy.
\end{aligned}
$$

For each fixed y, x_t, g_t, as $\zeta_{NT} \to 0$, for any $C > 0$, for all large N, T, $|\zeta_{NT}y| < C$. Recall $p_{\mathcal{Z}_t}(\cdot)$ is the pdf of $\mathcal{N}(0, \sigma_{h,x_t,g_t}^2)$ with $\sigma_{h,x_t,g_t}^2 := \text{plim}_{N\to\infty}\mathbb{E}[(h_t'\phi_0)^2|x_t, g_t, g_t'\phi_0 = 0]$. By

Assumption 4.4,

$$|p_{h_t'\phi_0|g_t,x_t,u_t=0}(\zeta_{NT}y)-p_{\mathcal{Z}_t}(0)| \leq \sup_{|z|<C} |p_{h_t'\phi_0|g_t,x_t,u_t=0}(z)-p_{\mathcal{Z}_t}(z)|+|p_{\mathcal{Z}_t}(\zeta_{NT}y)-p_{\mathcal{Z}_t}(0)| = o(1).$$

and $\sup_{x_t,g_t} p_{h_t'\phi_0|g_t,x_t,u_t=0}(\cdot) < C_0$ for some $C_0 > 0$ for all N,T. For each fixed g_t and all N,T, the integrand of $F_{NT,1}(g_t,x_t)$ is bounded by

$$|\left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y<0\right\} 1\left\{y>0\right\} p_{h_t'\phi_0|g_t,x_t,u_t=0}(\zeta_{NT}y)| \leq C_0|\left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y<0\right\} 1\left\{y>0\right\}|$$

with the right hand side being free of N,T and integrable with respect to y:

$$\int \left|\left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y<0\right\} 1\left\{y>0\right\}\right| dy = \frac{(g_t'\mathfrak{g})^2}{2} 1\{g_t'\mathfrak{g}<0\}.$$

Hence by the dominated convergence theorem, for each fixed g_t, x_t,

$$F_{NT,1}(g_t,x_t) \to F_1(g_t,x_t) := \int \left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y<0\right\} 1\left\{y>0\right\} p_{\mathcal{Z}_t}(0)dy = -\frac{1}{2}p_{\mathcal{Z}_t}(0)(g_t'\mathfrak{g})^2 1\{g_t'\mathfrak{g}<0\}.$$

Nnote that $p_{\mathcal{Z}_t}(0)$ does not depend on N,T, and is a function of x_t, g_t through σ^2_{h,x_t,g_t}. In addition, let $\mathcal{R}(x_t,g_t) = C_0(x_t'd_0)^2\frac{(g_t'\mathfrak{g})^2}{2}1\{g_t'\mathfrak{g}<0\}$. Then for all N,T,

$$\begin{aligned}
|(x_t'd_0)^2 F_{NT,1}(g_t,x_t)| &\leq (x_t'd_0)^2|\int \left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y<0\right\} 1\left\{y>0\right\} p_{h_t'\phi_0|g_t,x_t,u_t=0}(\zeta_{NT}y)dy| \\
&\leq C_0(x_t'd_0)^2 \int |\left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y<0\right\} 1\left\{y>0\right\}|dy \\
&= C_0(x_t'd_0)^2\frac{(g_t'\mathfrak{g})^2}{2}1\{g_t'\mathfrak{g}<0\} = \mathcal{R}(x_t,g_t)
\end{aligned}$$

Here $\mathcal{R}(x_t,g_t)$ is free of N,T, and $\mathbb{E}(|\mathcal{R}(x_t,g_t)||u_t=0) < \infty$. Therefore, still by the dominated convergence theorem, $\mathbb{E}[(x_t'd_0)^2F_{NT,1}(g_t,x_t)|u_t=0] \to \mathbb{E}[(x_t'd_0)^2F_1(g_t,x_t)|u_t=0]$. Using the similar argument, we also reach: $\mathbb{E}[(x_t'd_0)^2F_{NT,2}(g_t,x_t)|u_t=0] \to \mathbb{E}[(x_t'd_0)^2F_2(g_t,x_t)|u_t=0]$, where

$$F_2(g_t,x_t) := \int \left(g_t'\mathfrak{g}+y\right) 1\left\{g_t'\mathfrak{g}+y>0\right\} 1\left\{y\leq 0\right\} p_{\mathcal{Z}_t}(0)dy = \frac{1}{2}p_{\mathcal{Z}_t}(0)(g_t'\mathfrak{g})^2 1\{g_t'\mathfrak{g}>0\}.$$

So

$$\begin{aligned}
\breve{\mathbb{C}}_{NT,2}(\mathfrak{g}) &= -\widetilde{M}_{NT}2p_{u_t}(0)\mathbb{E}[(x_t'd_0)^2F_{NT,1}(g_t,x_t)|u_t=0] + \widetilde{M}_{NT}2p_{u_t}(0)\mathbb{E}[(x_t'd_0)^2F_{NT,2}(g_t,x_t)|u_t=0] \\
&\to -2\mathbb{E}[(x_t'd_0)^2p_{u_t}(0)F_1(g_t,x_t)|u_t=0] + 2\mathbb{E}[(x_t'd_0)^2p_{u_t}(0)F_2(g_t,x_t)|u_t=0] \\
&= (\mathbb{E}(x_t'd_0)^2(g_t'\mathfrak{g})^2|u_t=0,\mathcal{Z}_t=0)p_{u_t,\mathcal{Z}_t}(0,0) \\
&:= C(\mathfrak{g}).
\end{aligned}$$

Step II.2: $\breve{\mathbb{C}}_{NT,2}(H_T g) \xrightarrow{P} C(g)$

Again by the extended CMT (Lemma I.4), due to the pointwise convergence of $\breve{\mathbb{C}}_{NT,2}(\mathfrak{g})$, similar to the proof of step I.2, it suffices to prove, for any converging sequence $\mathfrak{g}_T \to \mathfrak{g}$ on a compact space, $|\breve{\mathbb{C}}_{NT,2}(\mathfrak{g}_T) - \breve{\mathbb{C}}_{NT,2}(\mathfrak{g})| \to 0$. By definition, $|\breve{\mathbb{C}}_{NT,2}(\mathfrak{g}_T) - \breve{\mathbb{C}}_{NT,2}(\mathfrak{g})| \leq a_1 + a_2$, where

$$\begin{aligned}
a_1 &= \widetilde{M}_{NT} 2 p_{u_t}(0) \mathbb{E}[(x_t' d_0)^2 |z(\mathfrak{g}_T) - z(\mathfrak{g})| | u_t = 0] \\
z(\mathfrak{g}_T) &:= \int \left(g_t' \mathfrak{g}_T + y\right) 1\left\{g_t' \mathfrak{g}_T + y < 0\right\} 1\{y > 0\} p_{h_t'\phi_0|g_t, x_t, u_t=0}(\zeta_{NT} y) dy \\
a_2 &= \widetilde{M}_{NT} 2 p_{u_t}(0) \mathbb{E}[(x_t' d_0)^2 |\widetilde{z}(\mathfrak{g}_T) - \widetilde{z}(\mathfrak{g})| | u_t = 0] \\
\widetilde{z}(\mathfrak{g}_T) &:= \int \left(g_t' \mathfrak{g}_T + y\right) 1\left\{g_t' \mathfrak{g}_T + y > 0\right\} 1\{y \leq 0\} p_{h_t'\phi_0|g_t, x_t, u_t=0}(\zeta_{NT} y) dy
\end{aligned}$$

and a_2 is defined similarly. Note that

$$\begin{aligned}
|z(\mathfrak{g}_T) - z(\mathfrak{g})| &\leq \int |\left(g_t' \mathfrak{g}_T + y\right) 1\left\{g_t' \mathfrak{g}_T + y < 0\right\} - \left(g_t' \mathfrak{g} + y\right) 1\left\{g_t' \mathfrak{g} + y < 0\right\} | 1\{y > 0\} \\
&\quad \cdot p_{h_t'\phi_0|g_t, x_t, u_t=0}(\zeta_{NT} y) dy \\
&\leq \int |g_t'(\mathfrak{g}_T - \mathfrak{g})| 1\left\{g_t' \mathfrak{g}_T + y < 0\right\} 1\{y > 0\} p_{h_t'\phi_0|g_t, x_t, u_t=0}(\zeta_{NT} y) dy \\
&\quad + \int |\left(g_t' \mathfrak{g} + y\right)| |1\left\{g_t' \mathfrak{g} + y < 0\right\} - 1\left\{g_t' \mathfrak{g}_T + y < 0\right\}| 1\{y > 0\} \\
&\quad \cdot p_{h_t'\phi_0|g_t, x_t, u_t=0}(\zeta_{NT} y) dy \\
&\leq C |g_t|_2^2 |\mathfrak{g}_T - \mathfrak{g}|_2.
\end{aligned}$$

Thus $a_1 \leq O(1) \mathbb{E}[(x_t' d_0)^2 |g_t|_2^2 | u_t = 0] |\mathfrak{g}_T - \mathfrak{g}|_2 = o(1)$. Similarly, $a_2 = o(1)$, implying $\breve{\mathbb{C}}_{NT,2}(\mathfrak{g}_T) \to \breve{\mathbb{C}}_{NT,2}(\mathfrak{g})$. Hence by the extended CMT, $\breve{\mathbb{C}}_{NT,2}(H_T g) \xrightarrow{P} C(g)$. So

$$\begin{aligned}
& l_{NT} \mathbb{E}_{|u_t=0} (x_t' \delta_0)^2 (A_1 - A_3 + A_2 - A_4) = \breve{\mathbb{C}}_{NT,2}(H_T g) + o(1) \\
\xrightarrow{P}\ & (\mathbb{E}(x_t' d_0)^2 ((g_t' H g)^2 | u_t = 0, \mathcal{Z}_t = 0) p_{u_t, \mathcal{Z}_t}(0, 0) := C(g).
\end{aligned}$$

Step II.3: $C(g) = \lim_{\omega \to 0} A(\omega, g)$

As $\omega \to 0$, we have that $\zeta_\omega = \omega^{1/3}$, $M_\omega = \omega^{-1/3}$. Still use (G.27) with $g_t' H g = a$, $\zeta_\omega^{-1} \mathcal{Z}_t = b$, and the formula $|a + b| - |b| = \Xi(a, b)$:

$$\begin{aligned}
A(\omega, g) &:= M_\omega \mathbb{E}\left[(x d_0)^2 \left(\left|g_t' H g + \zeta_\omega^{-1} \mathcal{Z}_t\right| - \left|\zeta_\omega^{-1} \mathcal{Z}_t\right|\right) \middle| u_t = 0\right] p_{u_t}(0) \\
&= M_\omega \mathbb{E}_{|u_t=0} (x_t' d_0)^2 p_{u_t}(0) \left(|a + b| - |b|\right) \\
&= -M_\omega \mathbb{E}_{|u_t=0} (x_t' d_0)^2 p_{u_t}(0) a 1\{a \leq 0\} 1\{b \leq 0\} \\
&\quad + M_\omega \mathbb{E}_{|u_t=0} (x_t' d_0)^2 p_{u_t}(0) a 1\{a > 0\} 1\{b > 0\} \\
&\quad + M_\omega \mathbb{E}_{|u_t=0} (x_t' d_0)^2 p_{u_t}(0) \Delta(a, b)
\end{aligned}$$

where $\Delta(a,b)$ denotes the sum of the other terms in the expression of $\Xi(a,b)$ given in (G.27). We now aim to obtain alternative expressions for the first two terms on the right hand side. Note that conditional on $(x_t, g_t, u_t = 0)$, $b = \zeta_\omega^{-1}\mathcal{Z}_t$ is Gaussian with zero mean, so the first term on the right hand side can be replaced with

$$\begin{aligned}
& -M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_t}(0) a 1\{a \le 0\} 1\{b \le 0\} \\
= \; & -M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_t}(0) a 1\{a \le 0\} 1\{b > 0\} \\
= \; & -M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_t}(0) a 1\{a \le 0\} 1\{b > -a\} - M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_t}(0) a 1\{a \le 0\} 1\{-a > b > 0\}
\end{aligned}$$

Similarly, $1\{b > 0\}$ in the second term on the right hand side of $A(\omega, g)$ can be replaced with

$$M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_t}(0) a 1\{a > 0\} 1\{b < -a\} + M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_t}(0) a 1\{a > 0\} 1\{-a < b < 0\}.$$

These alternative expressions can be combined with $\Delta(a,b)$, to reach: (note that $M_\omega = \zeta_k^{-1}$ and $\zeta_k \to 0$ as $k \to 0$),

$$\begin{aligned}
A(\omega, g) \;\; = \;\; & -2\zeta_k^{-1} \mathbb{E}_{|u_t=0} (x_t'd_0)^2 p_{u_t}(0) (a+b) 1\{a+b<0\} 1\{b>0\} \\
& +2\zeta_k^{-1} \mathbb{E}_{|u_t=0} (x_t'd_0)^2 p_{u_t}(0) (a+b) 1\{a+b>0\} 1\{b \le 0\} \\
= \;\; & -2\mathbb{E}_{|u_t=0} (x_t'd_0)^2 p_{u_t}(0) \int (a+b) 1\{a+b<0\} 1\{b>0\} p_{\mathcal{Z}_t}(\zeta_k b) db \\
& +2\mathbb{E}_{|u_t=0} (x_t'd_0)^2 p_{u_t}(0) \int (a+b) 1\{a+b>0\} 1\{b \le 0\} p_{\mathcal{Z}_t}(\zeta_k b) db \\
\to^{(1)} \;\; & -2\mathbb{E}_{|u_t=0} (x_t'd_0)^2 p_{u_t}(0) \int (a+b) 1\{a+b<0\} 1\{b>0\} p_{\mathcal{Z}_t}(0) db \\
& +2\mathbb{E}_{|u_t=0} (x_t'd_0)^2 p_{u_t}(0) \int (a+b) 1\{a+b>0\} 1\{b \le 0\} p_{\mathcal{Z}_t}(0) db \\
= \;\; & \mathbb{E}_{|u_t=0} \left(x_t'd_0\right)^2 p_{u_t}(0) p_{\mathcal{Z}_t}(0) a^2 \\
= \;\; & (\mathbb{E}(x_t'd_0)^2 (g_t'Hg)^2 | u_t = 0, \mathcal{Z}_t = 0) p_{u_t, \mathcal{Z}_t}(0,0) := C(g).
\end{aligned}$$

It remains to argue that (1) in the above limit holds by applying the DCT. First, for each fixed b, $p_{\mathcal{Z}_t}(\zeta_\omega b) \to p_{\mathcal{Z}_t}(0)$. Secondly, $\sup_x p_{\mathcal{Z}_t}(x) = \sup_x \frac{1}{\sqrt{2\pi\sigma^2_{h,x_t,g_t}}} \exp(-\frac{x^2}{2\sigma^2_{h,x_t,g_t}}) = (2\pi\sigma^2_{h,x_t,g_t})^{-1/2} < C_0$ for some $C_0 > 0$, due to $\inf_{x_t,g_t} \sigma^2_{h,x_t,g_t} > c_0$ (by the assumption). So in the integration: $(a = g_t'Hg)$

$$\mathcal{E}_{NT}(a) := \int (a+b) 1\{a+b<0\} 1\{b>0\} p_{\mathcal{Z}_t}(\zeta_k b) db,$$

$|(a+b) 1\{a+b<0\} 1\{b>0\} p_{\mathcal{Z}_t}(\zeta_k b)| < |(a+b) 1\{a+b<0\} 1\{b>0\}| C_0$, where the right hand side is free of N, T and is integrable: $\int |(a+b) 1\{a+b<0\} 1\{b>0\}| db < \infty$ for each fixed a. Then DCT implies $\mathcal{E}_{NT}(a) \to \mathcal{E}(a) := \int (a+b) 1\{a+b<0\} 1\{b>0\} p_{\mathcal{Z}_t}(0) db$ for each fixed a. Thirdly,

$$|(x_t^2 d_0)^2 \mathcal{E}_{NT}(a)| \leq (x_t^2 d_0)^2 C_0 \int |(a+b) 1\{a+b<0\} 1\{b>0\}| db \leq 0.5(x_t^2 d_0)^2 C_0 a^2$$

with $a = g_t' H g$, so that $0.5(x_t^2 d_0)^2 C_0 a^2$ is free of N, T and is integrable: $\mathbb{E}_{|u_t=0} 0.5(x_t^2 d_0)^2 C_0 a^2 < \infty$. Also, $(x_t^2 d_0)^2 \mathcal{E}_{NT}(a) \to (x_t^2 d_0)^2 \mathcal{E}(a)$ for each fixed x_t, g_t. Thus applying DCT again yields

$$\mathbb{E}_{|u_t=0}(x_t^2 d_0)^2 \mathcal{E}_{NT}(a) \to \mathbb{E}_{|u_t=0}(x_t^2 d_0)^2 \mathcal{E}(a).$$

The same argument also applies to the second term on the right hand side of (1).

H Proofs for Section 5

H.1 Proof of Theorem 5.1: known factor case

H.1.1 Proof of the distribution of LR

Below we prove, under $H_0 : h(\gamma_0) = 0$,

$$T \cdot LR \to^d \sigma_\varepsilon^{-2} \min_{g_h' \nabla h = 0} \mathbb{Q}(\infty, g_h) - \sigma_\varepsilon^{-2} \min_g \mathbb{Q}(\infty, g).$$

Proof. Define $\widehat{\gamma}_h = \arg\min_{\alpha, h(\gamma)=0} \mathbb{S}_T(\alpha, \gamma)$, $\widehat{\alpha}(\gamma) = \arg\min_\alpha \mathbb{S}_T(\alpha, \gamma)$, and $\widehat{\alpha}_h = \widehat{\alpha}(\widehat{\gamma}_h)$. Then,

$$\begin{aligned} T \min_{\alpha,\gamma} \mathbb{S}_T(\alpha,\gamma) LR &= T[\mathbb{S}_T(\widehat{\alpha}_h, \widehat{\gamma}_h) - \mathbb{S}_T(\widehat{\alpha}, \widehat{\gamma})] \\ &= A_1 + A_2 - A_3, \quad \text{where} \\ A_1 &= T[\mathbb{S}_T(\widehat{\alpha}_h, \widehat{\gamma}_h) - \mathbb{S}_T(\widehat{\alpha}_h, \gamma_0)] \\ A_2 &= T[\mathbb{S}_T(\widehat{\alpha}_h, \gamma_0) - \mathbb{S}_T(\widehat{\alpha}, \gamma_0)] \\ A_3 &= T[\mathbb{S}_T(\widehat{\alpha}, \widehat{\gamma}) - \mathbb{S}_T(\widehat{\alpha}, \gamma_0)]. \end{aligned}$$

Let us first prove a useful equality. Note that

$$T[\mathbb{S}_T(\alpha,\gamma) - \mathbb{S}_T(\alpha,\gamma_0)] = T[\mathbb{R}_T(\alpha,\gamma) - \mathbb{R}_T(\alpha,\gamma_0)] - T[\mathbb{G}_T(\alpha,\gamma) - \mathbb{G}_T(\alpha,\gamma_0)],$$

where $\mathbb{R}_T$ and $\mathbb{G}_T$ are defined in the proof of Lemma F.1. Also recall

$$\begin{aligned} \mathbb{K}_{2T}(g) &= T \cdot \mathbb{E}\left(x_t'\delta_0\right)^2 \left|1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t(\gamma_0)\right| \\ \mathbb{K}_{3T}(g) &= -2\sum_{t=1}^T \varepsilon_t x_t'\delta_0 \left(1_t\left(\gamma_0 + g \cdot r_T^{-1}\right) - 1_t(\gamma_0)\right). \end{aligned}$$

Here $r_T = T^{1-2\varphi}$, $1_t(\gamma) = 1\{f_t'\gamma > 0\}$. Uniformly over $|\gamma - \gamma_0|_2 < Cr_T^{-1}$, $|\alpha - \alpha_0|_2 < CT^{-1/2}$, and $g = r_T(\gamma - \gamma_0)$, we have

$$\begin{aligned}
& T[\mathbb{R}_T(\alpha,\gamma) - \mathbb{R}_T(\alpha,\gamma_0)] \\
= \ & T\delta' \frac{1}{T}\sum_t \left[x_t x_t' |1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}| - \mathbb{E}x_t x_t'|1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}|\right]\delta \\
& + T\alpha' \frac{2}{T}\sum_t [Z_t(\gamma) - Z_t(\gamma_0)]Z_t(\gamma_0)'(\alpha - \alpha_0) \\
& + T\mathbb{E}[(x_t'\delta)^2 - (x_t'\delta_0)^2]|1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}| + T\mathbb{E}(x_t'\delta_0)^2|1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}| \\
= \ & T\mathbb{E}(x_t'\delta_0)^2|1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}| + o_P(1) \\
= \ & \mathbb{K}_{2T}(g) + o_P(1)
\end{aligned}$$

and

$$\begin{aligned}
-T[\mathbb{G}_T(\alpha,\gamma) - \mathbb{G}_T(\alpha,\gamma_0)] &= -2\sum_{t=1}^T \varepsilon_t x_t'(\delta - \delta_0)(1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}) \\
& \quad -2\sum_{t=1}^T \varepsilon_t x_t'\delta_0(1\{f_t'\gamma > 0\} - 1\{f_t'\gamma_0 > 0\}) \\
&= \mathbb{K}_{3T}(g) + o_P(1).
\end{aligned}$$

Hence uniformly over $|\gamma - \gamma_0|_2 < Cr_T^{-1}$, $|\alpha - \alpha_0|_2 < CT^{-1/2}$, and $g = r_T(\gamma - \gamma_0)$,

$$T[\mathbb{S}_T(\alpha,\gamma) - \mathbb{S}_T(\alpha,\gamma_0)] = \mathbb{K}_{2T}(g) + \mathbb{K}_{3T}(g) + o_P(1). \tag{H.1}$$

We are now ready to analyze A_1. By Lemma H.1, $|\widehat{\gamma}_h - \gamma_0|_2 = O_P(T^{-(1-2\varphi)})$ under H_0. Also, in the proof of Lemma H.1 we have shown that $|\widehat{\alpha}_h - \alpha_0|_2 = O_P(T^{-1/2})$. Hence apply (H.1) with $\alpha = \widehat{\alpha}_h$ and $\gamma = \widehat{\gamma}_h$,

$$A_1 = T[\mathbb{S}_T(\widehat{\alpha}_h,\widehat{\gamma}_h) - \mathbb{S}_T(\widehat{\alpha}_h,\gamma_0)] = \mathbb{K}_{2T}(\widehat{g}_h) + \mathbb{K}_{3T}(\widehat{g}_h) + o_P(1).$$

To analyze the right hand side, recall that in Proof of Theorem 3.1,

$$\begin{aligned}
\mathbb{K}_T(a,g) &= T\left(\mathbb{S}_T\left(\alpha_0 + a\cdot T^{-1/2}, \gamma_0 + g\cdot r_T^{-1}\right) - \mathbb{S}_T(\alpha_0,\gamma_0)\right) \\
&= \mathbb{K}_{1T}(a) + \mathbb{K}_{2T}(g) + \mathbb{K}_{3T}(g) + o_P(1)
\end{aligned}$$

where $o_P(1)$ is uniform over any compact set. Define

$$\begin{aligned}
(\widehat{a}_h, \widehat{g}_h) &= \arg \min_{a, h(\gamma_0 + g_h r_T^{-1}) = 0} \mathbb{K}_T(a, g_h) \\
\widehat{g}_h &= T^{-1+2\varphi}(\widehat{\gamma}_h - \gamma_0)
\end{aligned}$$

$$\widetilde{g}_h = \arg \min_{h(\gamma_0+g_h r_T^{-1})=0} \mathbb{K}_{2T}(g_h) + \mathbb{K}_{3T}(g_h).$$

Then $\mathbb{K}_T(\widehat{a}_h, \widehat{g}_h) \leq \mathbb{K}_T(\widehat{a}_h, \widetilde{g}_h)$, implying

$$\begin{aligned}
\mathbb{K}_T(\widehat{a}_h, \widehat{g}_h) &= \mathbb{K}_{2T}(\widehat{g}_h) + \mathbb{K}_{3T}(\widehat{g}_h) + \mathbb{K}_{1T}(\widehat{a}_h) + o_P(1) \\
&\leq \mathbb{K}_T(\widehat{a}_h, \widetilde{g}_h) \\
\mathbb{K}_T(\widehat{a}_h, \widetilde{g}_h) &= \mathbb{K}_{2T}(\widetilde{g}_h) + \mathbb{K}_{3T}(\widetilde{g}_h) + \mathbb{K}_{1T}(\widehat{a}_h) + o_P(1) \\
&\leq \mathbb{K}_{2T}(\widehat{g}_h) + \mathbb{K}_{3T}(\widehat{g}_h) + \mathbb{K}_{1T}(\widehat{a}_h) + o_P(1).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{K}_{2T}(\widehat{g}_h) + \mathbb{K}_{3T}(\widehat{g}_h) &= \mathbb{K}_{2T}(\widetilde{g}_h) + \mathbb{K}_{3T}(\widetilde{g}_h) + o_P(1) \\
&= \min_{h(\gamma_0+g_h r_T^{-1})=0} \mathbb{K}_{2T}(g_h) + \mathbb{K}_{3T}(g_h) + o_P(1)
\end{aligned}$$

These imply, with $\mathbb{Q}_T(g) := \mathbb{K}_{2T}(g) + \mathbb{K}_{3T}(g)$,

$$\begin{aligned}
A_1 &= \mathbb{K}_{2T}(\widehat{g}_h) + \mathbb{K}_{3T}(\widehat{g}_h) + o_P(1) = \min_{h(\gamma_0+g_h r_T^{-1})=0} \mathbb{K}_{2T}(g_h) + \mathbb{K}_{3T}(g_h) + o_P(1) \\
&= \min_{h(\gamma_0+g_h r_T^{-1})=0} \mathbb{Q}_T(g_h) + o_P(1) \\
&= \min_{r_T\{h(\gamma_0+g_h r_T^{-1})-h(\gamma_0)\}=0} \mathbb{Q}_T(g_h) + o_P(1), \quad (\text{under } H_0 : h(\gamma_0) = 0), \\
&= \min_{g_h' \nabla h=0} \mathbb{Q}_T(g_h) + o_P(1).
\end{aligned}$$

As for A_2, Lemma H.1 shows that $A_2 = o_P(1)$. As for A_3, by definition $(\widehat{a}, \widehat{g}) = \arg\min_{a,g} \mathbb{K}_T(a, g)$ and $\widehat{g} = T^{-1+2\varphi}(\widehat{\gamma} - \gamma_0)$. Apply (H.1) with $\alpha = \widehat{\alpha}$ and $\gamma = \widehat{\gamma}$,

$$\begin{aligned}
A_3 &= T[\mathbb{S}_T(\widehat{\alpha}, \widehat{\gamma}) - \mathbb{S}_T(\widehat{\alpha}, \gamma_0)] = \mathbb{K}_{2T}(\widehat{g}) + \mathbb{K}_{3T}(\widehat{g}) + o_P(1) \\
&= \min_g \mathbb{Q}_T(g) + o_P(1).
\end{aligned}$$

Together, we have

$$T \min_{\alpha,\gamma} \mathbb{S}_T(\alpha, \gamma) LR = A_1 + A_2 - A_3 = \min_{g_h' \nabla h=0} \mathbb{Q}_T(g_h) - \min_g \mathbb{Q}_T(g) + o_P(1).$$

Note that $\mathbb{Q}_T(\cdot) \Rightarrow \mathbb{Q}(\infty, \cdot)$. In addition, the operator $\mathcal{P} : f \to \min_{g_h' \nabla h=0} f(g_h) - \min_g f(g)$ is continuous in f with respect to the metric (*essential supremum*)$|f_1 - f_2|_\infty = \inf\{M : |f_1(x) - f_2(x)| < M$ *almost surely*$\}$. Hence by the continuous mapping theorem, and the fact that $\min_{\alpha,\gamma} \mathbb{S}_T(\alpha, \gamma) \to^P \sigma_\varepsilon^2$,

$$T \cdot LR \to^d \sigma_\varepsilon^{-2} \min_{g_h' \nabla h=0} \mathbb{Q}(\infty, g_h) - \sigma_\varepsilon^{-2} \min_g \mathbb{Q}(\infty, g).$$

∎

H.1.2 Proof of the distribution of LR_k^*

Proof. We first prove that under $\mathcal{H}_0 : h(\gamma_0) = 0$,

$$T\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*)LR_k^* = \min_{g_h' \nabla h = 0} \mathbb{Q}_T^*(g_h) - \min_g \mathbb{Q}_T^*(g) + o_{P^*}(1)$$

where $\mathbb{Q}_T^*(g) = \sum_t \left(x_t'\widehat{\delta}\right)^2 \left|1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right| - 2\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t x_t'\widehat{\delta}\left(1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right).$

To do so, define

$$\begin{aligned}
\alpha^*(\gamma) &= \arg\min_\alpha \mathbb{S}_T^*(\alpha, \gamma), \\
\gamma^*(\alpha) &= \arg\min_\gamma \mathbb{S}_T^*(\alpha, \gamma) \\
\gamma_h^*(\alpha) &= \arg\min_{h(\gamma)=h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha, \gamma).
\end{aligned}$$

We have $T\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*)LR_k^* = T[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) - \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*)] = A_1^* + A_2^* - A_3^*$, where

$$A_1^* = T[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) - \mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma})], \quad A_2^* = T[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}) - \mathbb{S}_T(\widehat{\alpha}^*, \widehat{\gamma})], \quad A_3^* = T[\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) - \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma})].$$

Define

$$\begin{aligned}
\mathbb{R}_T^*(\alpha, \gamma) &:= \frac{1}{T}\sum_{t=1}^T \left(Z_t(\gamma)'\alpha - Z_t(\widehat{\gamma})'\widehat{\alpha}\right)^2 \\
\mathbb{G}_T^*(\alpha, \gamma) &:= \frac{2}{T}\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t Z_t(\gamma)'\alpha \\
\mathbb{K}_{1T}^*(a) &:= a'\frac{1}{T}\sum_t Z_t(\widehat{\gamma}) Z_t(\widehat{\gamma})' a - \frac{2}{\sqrt{T}}\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t Z_t(\widehat{\gamma})' a, \\
\mathbb{K}_{2T}^*(g) &:= T \cdot \frac{1}{T}\sum_t \left(x_t'\widehat{\delta}\right)^2 \left|1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right|, \\
\mathbb{K}_{3T}^*(g) &:= -2\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t x_t'\widehat{\delta}\left(1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right) \\
\mathbb{K}_T^*(a, g) &:= T\left(\mathbb{S}_T^*\left(\widehat{\alpha} + a \cdot T^{-1/2}, \widehat{\gamma} + g \cdot r_T^{-1}\right) - \mathbb{S}_T^*(\widehat{\alpha}, \widehat{\gamma})\right).
\end{aligned}$$

We first show two important equalities:
(i) $T[\mathbb{S}_T^*(\alpha, \gamma) - \mathbb{S}_T^*(\alpha, \widehat{\gamma})] = \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g) + o_{P^*}(1)$
(ii) $\mathbb{K}_T^*(a, g) = \mathbb{K}_{1T}^*(a) + \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g) + o_{P^*}(1)$,
where $o_{p^*}(1)$ is uniform over any compact set.

For (i), note that $T[\mathbb{S}_T^*(\alpha, \gamma) - \mathbb{S}_T^*(\alpha, \widehat{\gamma})] = T[\mathbb{R}_T^*(\alpha, \gamma) - \mathbb{R}_T^*(\alpha, \widehat{\gamma})] - T[\mathbb{G}_T^*(\alpha, \gamma) - \mathbb{G}_T^*(\alpha, \widehat{\gamma})]$.

To bound the right hand side, note that uniformly for $|\alpha - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$, $|\gamma - \widehat{\gamma}|_2 = O_{P^*}(r_T^{-1})$ and $g = r_T(\gamma - \widehat{\gamma})$,

$$\begin{aligned}
& T[\mathbb{R}_T^*(\alpha,\gamma) - \mathbb{R}_T^*(\alpha,\widehat{\gamma})] \\
= \; & T\frac{1}{T}\sum_t [\delta' x_t]^2 |1\{f_t'\gamma > 0\} - 1\{f_t'\widehat{\gamma} > 0\}| + T\frac{2}{T}\sum_t \delta' x_t (1\{f_t'\gamma > 0\} - 1\{f_t'\widehat{\gamma} > 0\}) Z_t(\widehat{\gamma})'(\alpha - \widehat{\alpha}) \\
= \; & \mathbb{K}_{2T}^*(g) + O_P(1) T^{1-\varphi} |\delta - \widehat{\delta}|_2 \frac{1}{T}\sum_t |x_t|_2^2 |1\{f_t'\gamma > 0\} - 1\{f_t'\widehat{\gamma} > 0\}| \\
= \; & \mathbb{K}_{2T}^*(g) + O_P(1) T^{1-\varphi} |\delta - \widehat{\delta}|_2 |\gamma - \widehat{\gamma}|_2 + o_P(1) = \mathbb{K}_{2T}^*(g) + o_{P^*}(1) \\
& -T[\mathbb{G}_T^*(\alpha,\gamma) - \mathbb{G}_T^*(\alpha,\widehat{\gamma})] = -2\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t x_t'(\delta - \widehat{\delta})(1\{f_t'\gamma > 0\} - 1\{f_t'\widehat{\gamma} > 0\}) + \mathbb{K}_{3T}^*(g) \\
= \; & \mathbb{K}_{3T}^*(g) + o_{P^*}(1),
\end{aligned}$$

where we applied Lemma F.2 on the bootstrap sampling space to show $\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t x_t'(\delta - \widehat{\delta})(1\{f_t'\gamma > 0\} - 1\{f_t'\widehat{\gamma} > 0\}) = o_{P^*}(1)$. Therefore, uniformly in g, $|\gamma - \widehat{\gamma}|_2 = O_{P^*}(r_T^{-1})$, and $|\alpha - \widehat{\alpha}|_2 = O_P(T^{-1/2})$,

$$T[\mathbb{S}_T^*(\alpha,\gamma) - \mathbb{S}_T^*(\alpha,\widehat{\gamma})] = \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g) + o_{P^*}(1). \tag{H.2}$$

For (ii), note that uniformly for $\alpha - \widehat{\alpha} = T^{-1/2}a$ and $\gamma - \widehat{\gamma} = r_T^{-1}g$, we have

$$\mathbb{K}_T^*(a,g) = \mathbb{K}_{1T}^*(a) + \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g) + \Delta_1^* + \Delta_2^* + \Delta_3^*$$

where

$$\begin{aligned}
\Delta_1^*(\alpha,\gamma) &= 2\sum_t x_t'(\widehat{\delta} - \delta)\eta_t \widehat{\epsilon}_t (1_t(\gamma) - 1_t(\widehat{\gamma})) = o_{P^*}(1) \\
\Delta_2^*(\alpha,\gamma) &= \frac{2}{\sqrt{T}}\sum_t a' Z_t(\widehat{\gamma}) x_t'\delta (1_t(\gamma) - 1_t(\widehat{\gamma})) \\
\Delta_3^*(\alpha,\gamma) &= o_P(1)\sum_t [(x_t'\delta)^2 - (x_t'\widehat{\delta})^2] |1_t(\gamma) - 1_t(\widehat{\gamma})| = o_P(1)
\end{aligned}$$

where we applied Lemma F.2 on the bootstrap sampling space to bound the first term, and applied the same lemma on the original space to bound the other two terms.

We are now ready to analyze A_1^*. By Lemma H.2, $|\widehat{\alpha}_h^* - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$, and $|\widehat{\gamma}_h^* - \widehat{\gamma}|_2 = O_P(T^{-(1-2\varphi)})$. Apply (H.2) with $\alpha = \widehat{\alpha}_h^*$ and $\gamma = \widehat{\gamma}_h^* = \widehat{\gamma} + \widehat{g}_h^* r_T^{-1}$,

$$A_1^* = T[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) - \mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma})] = \mathbb{K}_{2T}^*(\widehat{g}_h^*) + \mathbb{K}_{3T}^*(\widehat{g}_h^*) + o_{P^*}(1).$$

Define

$$\widehat{a}^* = \sqrt{T}(\widehat{\alpha}_h^* - \widehat{\alpha}), \quad \widehat{g}_h^* = r_T(\widehat{\gamma}_h^* - \widehat{\gamma})$$

$$\widetilde{g}_h^* \;:=\; \arg\min_{h(\widehat{\gamma}+gr_T^{-1})=h(\widehat{\gamma})} \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g), \quad h(\widehat{\gamma} + \widetilde{g}_h^* r_T^{-1}) = h(\widehat{\gamma}).$$

By Lemma H.2, $\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) \leq \min_{\alpha, h(\gamma)=h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha, \gamma) + o_{P^*}(T^{-1})$, and $h(\widehat{\gamma}_h^*) = h(\widehat{\gamma})$ we have

$$\begin{aligned} \mathbb{K}_T^*(\widehat{a}^*, \widehat{g}_h^*) &= T\left(\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) - \mathbb{S}_T^*(\widehat{\alpha}, \widehat{\gamma})\right) \leq T\left(\min_{\alpha, h(\gamma)=h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha, \gamma) - \mathbb{S}_T^*(\widehat{\alpha}, \widehat{\gamma})\right) + o_{P^*}(1) \\ &= \min_{a, h(\widehat{\gamma}+r_T^{-1}g)=h(\widehat{\gamma})} \mathbb{K}_T^*(a, g) + o_{P^*}(1) \leq \mathbb{K}_T(\widehat{a}^*, \widetilde{g}_h^*) + o_{P^*}(1). \end{aligned}$$

So by $\mathbb{K}_T^*(a, g) = \mathbb{K}_{1T}^*(a) + \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g) + o_{P^*}(1)$,

$$\mathbb{K}_{2T}^*(\widehat{g}_h^*) + \mathbb{K}_{3T}^*(\widehat{g}_h^*) + \mathbb{K}_{1T}^*(\widehat{a}_h^*) + o_{P^*}(1) = \mathbb{K}_T^*(\widehat{a}^*, \widehat{g}_h^*) \leq \mathbb{K}_T^*(\widehat{a}^*, \widetilde{g}_h^*) + o_{P^*}(1).$$

On the other hand, by the definition of $\widetilde{g}_h^*$,

$$\begin{aligned} \mathbb{K}_T^*(\widehat{a}^*, \widetilde{g}_h^*) &= \mathbb{K}_{2T}^*(\widetilde{g}_h^*) + \mathbb{K}_{3T}^*(\widetilde{g}_h^*) + \mathbb{K}_{1T}^*(\widehat{a}_h^*) + o_{P^*}(1) \\ &\leq \mathbb{K}_{2T}^*(\widehat{g}_h^*) + \mathbb{K}_{3T}^*(\widehat{g}_h^*) + \mathbb{K}_{1T}^*(\widehat{a}_h^*) + o_{P^*}(1) \end{aligned}$$

and note that $\mathbb{Q}_T^*(g) = \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g)$. So

$$\begin{aligned} \mathbb{K}_{2T}^*(\widehat{g}_h^*) + \mathbb{K}_{3T}^*(\widehat{g}_h^*) &= \mathbb{K}_{2T}^*(\widetilde{g}_h^*) + \mathbb{K}_{3T}^*(\widetilde{g}_h^*) + o_{P^*}(1) \\ &= \min_{h(\widehat{\gamma}+gr_T^{-1})=h(\widehat{\gamma})} \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g) + o_{P^*}(1) \\ &= \min_{h(\widehat{\gamma}+gr_T^{-1})=h(\widehat{\gamma})} \mathbb{Q}_T^*(g) + o_{P^*}(1). \end{aligned}$$

These imply

$$A_1^* \;=\; \mathbb{K}_{2T}^*(\widehat{g}_h^*) + \mathbb{K}_{3T}^*(\widehat{g}_h^*) + o_{P^*}(1) = \min_{h(\widehat{\gamma}+gr_T^{-1})=h(\widehat{\gamma})} \mathbb{Q}_T^*(g) + o_{P^*}(1).$$

As for A_2^*, Lemma H.2 shows that $\widehat{\alpha}_h^* - \widehat{\alpha}^* = o_{P^*}(T^{-1/2})$. Hence similar proof as in Lemma H.1 shows $A_2^* = o_{P^*}(1)$.

As for A_3^*, let $\widehat{g}^* = r_T(\widehat{\gamma}^* - \widehat{\gamma})$. Apply (H.2) with $\alpha = \widehat{\alpha}^*$, and $\gamma = \widehat{\gamma}^*$, then $A_3^* = [\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) - \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma})] = \mathbb{K}_{2T}(\widehat{g}^*) + \mathbb{K}_{3T}(\widehat{g}^*) + o_{P^*}(1)$. Now let $\widetilde{g}^* = \arg\min_g \mathbb{K}_{2T}^*(g) + \mathbb{K}_{3T}^*(g)$. Then by Lemma H.2 , $\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) \leq \min_{\alpha,\gamma} \mathbb{S}_T^*(\alpha, \gamma) + o_{P^*}(T^{-1})$. Hence,

$$\begin{aligned} & \mathbb{K}_{1T}^*(\widehat{a}^*) + \mathbb{K}_{2T}^*(\widehat{g}^*) + \mathbb{K}_{3T}^*(\widehat{g}^*) + o_{P^*}(1) \\ =\; & \mathbb{K}_T^*(\widehat{a}^*, \widehat{g}^*) \\ =\; & T\left(\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) - \mathbb{S}_T^*(\widehat{\alpha}, \widehat{\gamma})\right) \leq T\left(\min_{\alpha,\gamma} \mathbb{S}_T^*(\alpha, \gamma) - \mathbb{S}_T^*(\widehat{\alpha}, \widehat{\gamma})\right) + o_{P^*}(1) \\ =\; & \min_{a,g} \mathbb{K}_T^*(a, g) + o_{P^*}(1) \leq \mathbb{K}_T^*(\widehat{a}^*, \widetilde{g}^*) + o_{P^*}(1) \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{K}^*_{1T}(\widehat{a}^*) + \mathbb{K}^*_{2T}(\widetilde{g}^*) + \mathbb{K}^*_{3T}(\widetilde{g}^*) + o_{P^*}(1) \\
&\leq \mathbb{K}^*_{1T}(\widehat{a}^*) + \mathbb{K}^*_{2T}(\widehat{g}^*) + \mathbb{K}^*_{3T}(\widehat{g}^*) + o_{P^*}(1).
\end{aligned}
$$

This implies $\mathbb{K}^*_{2T}(\widetilde{g}^*) + \mathbb{K}^*_{3T}(\widetilde{g}^*) \leq \mathbb{K}^*_{2T}(\widehat{g}^*) + \mathbb{K}^*_{3T}(\widehat{g}^*) \leq \mathbb{K}^*_{2T}(\widetilde{g}^*) + \mathbb{K}^*_{3T}(\widetilde{g}^*) + o_{P^*}(1)$. So

$$
\begin{aligned}
A^*_3 &= \mathbb{K}^*_{2T}(\widehat{g}^*) + \mathbb{K}^*_{3T}(\widehat{g}^*) + o_{P^*}(1) = \mathbb{K}^*_{2T}(\widetilde{g}^*) + \mathbb{K}^*_{3T}(\widetilde{g}^*) + o_{P^*}(1) \\
&= \min_g \mathbb{Q}^*_T(g) + o_{P^*}(1).
\end{aligned}
$$

Together, we have

$$
\begin{aligned}
T\mathbb{S}^*_T(\widehat{\alpha}^*, \widehat{\gamma}^*)LR^*_k &= A^*_1 + A^*_2 - A^*_3 = \min_{g_h : h(\widehat{\gamma} + g_h r_T^{-1}) = h(\widehat{\gamma})} \mathbb{Q}^*_T(g_h) - \min_g \mathbb{Q}^*_T(g) + o_{P^*}(1) \\
&= \min_{r_T\{h(\widehat{\gamma} + g_h r_T^{-1}) - h(\widehat{\gamma})\} = 0} \mathbb{Q}^*_T(g_h) - \min_g \mathbb{Q}^*_T(g) + o_{P^*}(1) \\
&= \min_{g_h' \nabla h = 0} \mathbb{Q}^*_T(g_h) - \min_g \mathbb{Q}^*_T(g) + o_{P^*}(1).
\end{aligned}
$$

Here ∇h is constant since h is linear.

Next, recall

$$
\begin{aligned}
\mathbb{K}^*_{2T}(g) &:= T \cdot \frac{1}{T} \sum_t \left(x_t' \widehat{\delta}\right)^2 \left|1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right| \\
&= T \cdot \frac{1}{T} \sum_t \left(x_t' \delta_0\right)^2 \left|1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right| + o_P(1) \\
&= M_T(\widehat{\gamma}, g) + o_P(1)
\end{aligned}
$$

where

$$
M_T(\gamma, g) = T \cdot \mathbb{E}\left(x_t' \delta_0\right)^2 \left|1_t\left(\gamma + g \cdot r_T^{-1}\right) - 1_t(\gamma)\right|.
$$

For any $\gamma_T \to \gamma_0$, and fixed g, we have $M_T(\gamma_T, g) \to \mathbb{Q}(\infty, g)$. It then follows from the extended continuous mapping theorem that $M_T(\widehat{\gamma}, g) \to^P \mathbb{Q}(\infty, g)$ for each g. So $\mathbb{K}^*_{2T} = \mathbb{Q}(\infty, g) + o_P(1)$ pointwise for each g.

Next, for $\mathbb{K}^*_{3T}(g) := -2\sum_{t=1}^T \eta_t \widehat{\varepsilon}_t x_t' \widehat{\delta}\left(1_t\left(\widehat{\gamma} + g \cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right)$, Lemma H.4 shows that in the known factor case, $\mathbb{K}^*_{3T}(g) \Rightarrow^* 2W(g)$.

So $\mathbb{Q}^*_T(.) = \mathbb{K}^*_{2T}(.) + \mathbb{K}^*_{3T}(.) \Rightarrow^* \mathbb{Q}(\infty, .)$. Here $\Rightarrow^*$ denotes the weak convergence with respect to the bootstrap distribution. It follows that

$$
\begin{aligned}
T\mathbb{S}^*_T(\widehat{\alpha}^*, \widehat{\gamma}^*)LR^*_k &= \min_{g_h' \nabla h = 0} \mathbb{Q}^*_T(g_h) - \min_g \mathbb{Q}^*_T(g) + o_{P^*}(1) \\
&\to^{d^*} \min_{g_h' \nabla h = 0} \mathbb{Q}(\infty, g_h) - \min_g \mathbb{Q}(\infty, g).
\end{aligned}
$$

In addition,

$$\begin{aligned}\mathbb{S}_T^*(\widehat{\alpha}^*,\widehat{\gamma}^*) &= \mathbb{S}_T^*(\widehat{\alpha},\widehat{\gamma})+o_{P^*}(1)=\frac{1}{T}\sum_t(\eta_t\widehat{\varepsilon}_t)^2+o_{P^*}(1)\\ &= \mathbb{E}^*\frac{1}{T}\sum_t(\eta_t\widehat{\varepsilon}_t)^2+o_{P^*}(1)=\frac{1}{T}\sum_t\widehat{\varepsilon}_t^2+o_{P^*}(1)=\sigma_\varepsilon^2+o_{P^*}(1).\end{aligned}$$

Thus $T\cdot LR_k^*\rightarrow^{d^*}\sigma_\varepsilon^{-2}\min_{g_h'\nabla h=0}\mathbb{Q}(\infty,g_h)-\sigma_\varepsilon^{-2}\min_g\mathbb{Q}(\infty,g)$.
∎

H.1.3 Technical Lemmas

Lemma H.1. *Under $\mathcal{H}_0$,*
(i) $|\widehat{\gamma}_h-\gamma_0|_2=O_P(T^{-(1-2\varphi)})$.
(ii) $T[\mathbb{S}_T(\widehat{\alpha}_h,\gamma_0)-\mathbb{S}_T(\widehat{\alpha},\gamma_0)]=o_P(1)$

Proof. (i) The proof is ver similar to that of the rate for $\widehat{\gamma}$, so we only briefly sketch the main steps. First of all, $h(\gamma_0)=0$ and $h(\widehat{\gamma}_h)=0$. By definition, we have

$$\begin{aligned}0 &\geq \mathbb{S}_T(\widehat{\alpha}_h,\widehat{\gamma}_h)-\mathbb{S}_T(\alpha_0,\gamma_0)\\ &= \mathbb{R}_T(\widehat{\alpha}_h,\widehat{\gamma}_h)-\mathbb{G}_T(\widehat{\alpha}_h,\widehat{\gamma}_h)+\mathbb{G}_T(\alpha_0,\gamma_0),\end{aligned}$$

Similarly as before, we can find some $c,c'>0$ such that for sufficiently small $|\alpha-\alpha_0|_2$

$$\begin{aligned}R(\alpha,\gamma) &= \mathbb{E}\left(Z_t(\gamma)'(\alpha-\alpha_0)\right)^2+\mathbb{E}\left(x_t'\delta_0\left(1_t(\gamma)-1_t(\gamma_0)\right)\right)^2\\ &\quad+2\mathbb{E}\left(x_t'\delta_0\left(1_t(\gamma)-1_t(\gamma_0)\right)\right)Z_t(\gamma)'(\alpha-\alpha_0)\\ &\geq c\,|\alpha-\alpha_0|_2^2+cT^{-2\varphi}\,|\gamma-\gamma_0|_2-c'\,|\alpha-\alpha_0|_2\,|\gamma-\gamma_0|_2\,T^{-\varphi},\end{aligned}$$

where the first inequality is from the bounds and the second from the condition that $|\alpha-\alpha_0|_2$ is small. (this is guaranteed since $\widehat{\alpha}_h$ is consistent under H_0) Furthermore, we still have, for $0<\eta<c$,

$$\begin{aligned}|\mathbb{G}_T(\alpha,\gamma)-\mathbb{G}_T(\alpha_0,\gamma_0)| &\leq O_P\left(\frac{1}{\sqrt{T}}\right)|\alpha-\alpha_0|_2+\eta T^{-2\varphi}\,|\gamma-\gamma_0|_2+O_P\left(\frac{1}{T}\right)\\ |\mathbb{R}_T(\alpha,\gamma)-R(\alpha,\gamma)| &\leq \eta\,|\alpha-\alpha_0|_2^2+\eta T^{-2\varphi}\,|\gamma-\gamma_0|_2+O_P\left(\frac{1}{T}\right),\end{aligned}$$

where the inequality is uniform in α and γ in the sense that the sequences $O_P(\cdot)$ and $o_P(\cdot)$ do not depend on α and γ. Since

$$R(\widehat{\alpha}_h,\widehat{\gamma}_h)\leq|\mathbb{G}_T(\widehat{\alpha}_h,\widehat{\gamma}_h)-\mathbb{G}_T(\alpha_0,\gamma_0)|+|\mathbb{R}_T(\widehat{\alpha}_h,\widehat{\gamma}_h)-R(\widehat{\alpha}_h,\widehat{\gamma}_h)|,$$

we conclude that

$$(c-\eta)\left(|\widehat{\alpha}_h-\alpha_0|_2^2+T^{-2\varphi}\,|\widehat{\gamma}_h-\gamma_0|_2\right)\le O_P\left(\frac{1}{\sqrt{T}}\right)|\widehat{\alpha}_h-\alpha_0|_2+O_P\left(\frac{1}{T}\right).$$

implying

$$|\widehat{\gamma}_h-\gamma_0|_2=O_P\left(\frac{1}{T^{1-2\varphi}}\right).$$

(ii) First we show that $|\widehat{\alpha}_h-\widehat{\alpha}|_2=o_P(T^{-1/2})$ under H_0. Let $\widehat{Z}_h=Z(\widehat{\gamma}_h)$. Straightforward calculations yield

$$\begin{aligned}\widehat{\alpha}_h-\widehat{\alpha} &= (\widehat{Z}_h'\widehat{Z}_h)^{-1}(\widehat{Z}_h-\widehat{Z})'(Z-\widehat{Z}_h)\alpha_0+(\widehat{Z}_h'\widehat{Z}_h)^{-1}\widehat{Z}'(\widehat{Z}-\widehat{Z}_h)\alpha_0+(\widehat{Z}_h'\widehat{Z}_h)^{-1}(\widehat{Z}_h-\widehat{Z})'\epsilon\\ &\quad+[(\widehat{Z}_h'\widehat{Z}_h)^{-1}-(\widehat{Z}'\widehat{Z})^{-1}][\widehat{Z}'(Z-\widehat{Z})\alpha_0+Z'\epsilon+(\widehat{Z}-Z)'\epsilon]\end{aligned}$$

which is $o_P(T^{-1/2})$ since $|\widehat{\gamma}_h-\gamma_0|_2=o(T^{-(0.5-\varphi)})$ under H_0. Then

$$\begin{aligned}&T[\mathbb{S}_T(\widehat{\alpha}_h,\gamma_0)-\mathbb{S}_T(\widehat{\alpha},\gamma_0)]\\ =\;&T(\widehat{\alpha}_h-\widehat{\alpha})'\frac{1}{T}\sum_t Z_t(\gamma_0)Z_t(\gamma_0)'(\widehat{\alpha}_h-\widehat{\alpha})+T(\widehat{\alpha}-\widehat{\alpha}_h)'\frac{2}{T}\sum_t Z_t(\gamma_0)\varepsilon_t\\ &+T\frac{2}{T}(\alpha_0-\widehat{\alpha})\sum_t Z_t(\gamma_0)Z_t(\gamma_0)'(\widehat{\alpha}-\widehat{\alpha}_h)\\ =\;&O_P(T)|\widehat{\alpha}_h-\widehat{\alpha}|_2^2+O_P(\sqrt{T})|\widehat{\alpha}_h-\widehat{\alpha}|_2=o_P(1).\end{aligned}$$

∎

Lemma H.2. *In the known factor case, the k-step bootstrap estimators $(\widehat{\alpha}^*,\widehat{\gamma}^*,\widehat{\gamma}_h^*)$ satisfy:*

$$\begin{aligned}\mathbb{S}_T^*(\widehat{\alpha}^*,\widehat{\gamma}^*) &\le \min_{\alpha,\gamma}\mathbb{S}_T^*(\alpha,\gamma)+o_{P^*}(T^{-1}).\\ \mathbb{S}_T^*(\widehat{\alpha}_h^*,\widehat{\gamma}_h^*) &\le \min_{\alpha,h(\gamma)=h(\widehat{\gamma})}\mathbb{S}_T^*(\alpha,\gamma)+o_{P^*}(T^{-1}),\quad h(\widehat{\gamma}_h^*)=h(\widehat{\gamma})\\ |\widehat{\alpha}^*-\widehat{\alpha}|_2 &= O_{P^*}(T^{-1/2}),\quad |\widehat{\alpha}_h^*-\widehat{\alpha}|_2=O_{P^*}(T^{-1/2}),\quad |\widehat{\alpha}^*-\widehat{\alpha}_h^*|_2=o_{P^*}(T^{-1/2})\\ |\widehat{\gamma}_h^*-\widehat{\gamma}|_2 &= O_P(T^{-(1-2\varphi)})\\ |\widehat{\gamma}^*-\widehat{\gamma}|_2 &= O_P(T^{-(1-2\varphi)}).\end{aligned}$$

Proof. Define

$$\begin{aligned}(\alpha_g^*,\gamma_g^*) &= \arg\min\mathbb{S}_T^*(\alpha,\gamma).\\ (\alpha_{g,h}^*,\gamma_{g,h}^*) &= \arg\min_{\alpha,h(\gamma)=h(\widehat{\gamma})}\mathbb{S}_T^*(\alpha,\gamma)\\ \alpha^*(\gamma) &= \arg\min_{\alpha}\mathbb{S}_T^*(\alpha,\gamma),\\ \gamma^*(\alpha) &= \arg\min_{\gamma}\mathbb{S}_T^*(\alpha,\gamma),\\ \gamma_h^*(\alpha) &= \arg\min_{\gamma:h(\gamma)=h(\widehat{\gamma})}\mathbb{S}_T^*(\alpha,\gamma).\end{aligned}$$

Our proof is divided into the following steps.

step 0: $|\gamma^*_{g,h}-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$, $|\gamma^*_g-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$ and $|\alpha^*_g-\widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$.

step 1: if $|\gamma-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$, then $|\alpha^*(\gamma)-\widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$.

step 2: in addition, $|\alpha^*(\gamma)-\alpha^*_g|_2 = o_{P^*}(T^{-1/2})$, and $|\alpha^*(\gamma)-\alpha^*_{g,h}|_2 = o_{P^*}(T^{-1/2})$.

step 3: if $|\alpha-\alpha^*_g|_2 = o_{P^*}(T^{-1/2})$, and $|\alpha-\alpha^*_{g,h}|_2 = o_{P^*}(T^{-1/2})$, then $\mathbb{S}^*_T(\alpha,\gamma^*(\alpha)) \le \min_{\alpha,\gamma}\mathbb{S}^*_T(\alpha,\gamma)+o_{P^*}(T^{-1})$, and $\mathbb{S}^*_T(\alpha,\gamma^*_h(\alpha)) \le \min_{\alpha,h(\gamma)=h(\widehat{\gamma})}\mathbb{S}^*_T(\alpha,\gamma)+o_{P^*}(T^{-1})$.

step 4: in addition, $|\gamma^*(\alpha)-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$ and $|\gamma^*_h(\alpha)-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$.

Once the above steps are successfully achieved, then the proof is completed by the following argument. Recall that $\widehat{\gamma}^{*,0}=\widehat{\gamma}^{*,0}_h=\widehat{\gamma}$. Also, for $l\ge 1$, $\widehat{\alpha}^{*,l}=\alpha^*(\widehat{\gamma}^{*,l-1})$, $\widehat{\alpha}^{*,l}_h=\alpha^*(\widehat{\gamma}^{*,l-1}_h)$, $\widehat{\gamma}^{*,l}=\gamma^*(\widehat{\alpha}^{*,l})$, $\widehat{\gamma}^{*,l}_h=\gamma^*_h(\widehat{\alpha}^{*,l}_h)$, and $\widehat{\alpha}^*=\widehat{\alpha}^{*,k}$, $\widehat{\gamma}^*=\widehat{\gamma}^{*,k}$, and $\widehat{\gamma}^*_h=\widehat{\gamma}^{*,k}_h$.

For $k=1$, $\widehat{\gamma}^{*,0}=\widehat{\gamma}^{*,0}_h=\widehat{\gamma}$. Hence by step 1, $|\widehat{\alpha}^{*,1}-\widehat{\alpha}|_2 = O_{P^*}(T^{-1/2}) = |\widehat{\alpha}^{*,1}_h-\widehat{\alpha}|_2$. Conditions of step 3 are satisfied due to step 2, hence for $\alpha=\alpha^*(\widehat{\gamma}^{*,0})$ in step 3,

$$\mathbb{S}^*_T(\widehat{\alpha}^{*,1},\widehat{\gamma}^{*,1}) \le \min_{\alpha,\gamma}\mathbb{S}^*_T(\alpha,\gamma)+o_{P^*}(T^{-1})$$

and for $\alpha=\alpha^*(\widehat{\gamma}^{*,0}_h)$ in step 3,

$$\mathbb{S}^*_T(\widehat{\alpha}^{*,1}_h,\widehat{\gamma}^{*,1}_h) \le \min_{\alpha,h(\gamma)=h(\widehat{\gamma})}\mathbb{S}^*_T(\alpha,\gamma)+o_{P^*}(T^{-1}).$$

By step 4, $|\widehat{\gamma}^{*,1}-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$ and $|\widehat{\gamma}^{*,1}_h-\widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$. Thus results of Lemma H.2 are verified if we stop after k step(s) for $k=1$.

For $k=2$, the previous step 4 ensures that we can apply step 1 respectively with $\gamma=\widehat{\gamma}^{*,1}$ and $\gamma=\widehat{\gamma}^{*,1}_h$. Thus the same argument yields Lemma H.2 is verified for $k=2$. We can employ the mathematical induction to conclude that Lemma H.2 is verified for all $k\ge 1$.

Proof of Step 0.

In the bootstrap world, $\widehat{\gamma}$ is the true value while γ^*_g is the least squares estimator. Also, by the definition of $(\alpha^*_{g,h},\gamma^*_{g,h})$, we have

$$\mathbb{S}^*_T(\alpha^*_{g,h},\gamma^*_{g,h}) \le \mathbb{S}^*_T(\widehat{\alpha},\widehat{\gamma}), \quad \mathbb{S}^*_T(\alpha^*_g,\gamma^*_g) \le \mathbb{S}^*_T(\widehat{\alpha},\widehat{\gamma}).$$

Hence the proof of this step is simply the bootstrap version of the proof of the rates of convergence in the original sampling space. We thus omit its proof to avoid repetitions.

Proof of Step 1.

For a generic γ, let $A(\gamma) := \frac{1}{T}\sum_{t=1}^{T} Z_t(\gamma)\, Z_t(\gamma)'$.

$$\alpha^*(\gamma)-\widehat{\alpha} \quad = \quad A(\gamma)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t(\gamma)\,\widehat{\varepsilon}_t\eta_t+\frac{1}{T}\sum_{t=1}^{T} Z_t(\gamma)\,x_t'\widehat{\delta}\left(1_t(\gamma)-1_t(\widehat{\gamma})\right)\right).$$

So conditional on the event $|\widehat{\gamma}-\gamma_0|_2 \leq CT^{-(1-2\varphi)}$ and uniformly in $|\gamma-\widehat{\gamma}| \leq CT^{-(1-2\varphi)}$,

$$\begin{aligned}
&\left|\alpha^*(\gamma)-\widehat{\alpha}-A(\widehat{\gamma})^{-1}\frac{1}{T}\sum_{t=1}^{T}Z_t(\widehat{\gamma})\widehat{\varepsilon}_t\eta_t\right| \\
\leq\ &\left|(A(\gamma)^{-1}-A(\widehat{\gamma})^{-1})\frac{1}{T}\sum_{t=1}^{T}Z_t(\widehat{\gamma})\widehat{\varepsilon}_t\eta_t\right| + |A(\gamma)^{-1}| \sup_{|\gamma-\widehat{\gamma}|_2\leq CT^{-(1-2\varphi)}}\left|\frac{1}{T}\sum_{t=1}^{T}[Z_t(\gamma)-Z_t(\widehat{\gamma})]\widehat{\varepsilon}_t\eta_t\right| \\
&+|A(\gamma)^{-1}O_P(T^{-\varphi})| \sup_{|\gamma-\gamma_0|_2\leq CT^{-(1-2\varphi)}}\left|\frac{1}{T}\sum_{t=1}^{T}|x_t|_2^2|1_t(\gamma)-1_t(\gamma_0)|-\mathbb{E}|x_t|_2^2|1_t(\gamma)-1_t(\gamma_0)|\right| \\
&+|A(\gamma)^{-1}O_P(T^{-\varphi}) \sup_{|\gamma-\gamma_0|_2\leq CT^{-(1-2\varphi)}}\mathbb{E}|x_t|_2^2|1_t(\gamma)-1_t(\gamma_0)| \\
=\ &o_{P^*}(T^{-1/2}).
\end{aligned}$$

Thus we have proved, uniformly over $|\gamma-\widehat{\gamma}| \leq CT^{-(1-2\varphi)}$,

$$\alpha^*(\gamma)-\widehat{\alpha} = A(\widehat{\gamma})^{-1})\frac{1}{T}\sum_{t=1}^{T}Z_t(\widehat{\gamma})\widehat{\varepsilon}_t\eta_t + o_{P^*}(T^{-1/2}). \tag{H.3}$$

Proof of Step 2.

Note that $\alpha_g^* = \alpha^*(\gamma_g^*)$ and $\alpha_{m,h}^* = \alpha^*(\gamma_{g,h}^*)$. Respectively letting $\gamma = \gamma_{g,h}^*$ and $\gamma = \gamma_g^*$ in (H.3) yields (by step 2)

$$\begin{aligned}
\alpha^*(\gamma)-\widehat{\alpha} &= A(\widehat{\gamma})^{-1})\frac{1}{T}\sum_t Z_t(\widehat{\gamma})\widehat{\varepsilon}_t\eta_t + o_{P^*}(T^{-1/2}) \\
\alpha_m^*-\widehat{\alpha} &= A(\widehat{\gamma})^{-1})\frac{1}{T}\sum_t Z_t(\widehat{\gamma})\widehat{\varepsilon}_t\eta_t + o_{P^*}(T^{-1/2}) \\
\alpha_{g,h}^*-\widehat{\alpha} &= A(\widehat{\gamma})^{-1})\frac{1}{T}\sum_t Z_t(\widehat{\gamma})\widehat{\varepsilon}_t\eta_t + o_{P^*}(T^{-1/2}).
\end{aligned}$$

Thus $\alpha^*(\gamma)-\alpha_g^* = o_{P^*}(T^{-1/2})$ and $\alpha^*(\gamma)-\alpha_{g,h}^* = o_{P^*}(T^{-1/2})$.

Proof of Step 3.

By the definition of $\gamma^*(\alpha)$ and $\gamma_h^*(\alpha)$,

$$\begin{aligned}
\mathbb{S}_T^*(\alpha,\gamma^*(\alpha)) &\leq \mathbb{S}_T^*(\alpha,\gamma_g^*) = \mathbb{S}_T^*(\alpha_g^*,\gamma_g^*) + \mathbb{S}_T^*(\alpha,\gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*,\gamma_g^*) \\
\mathbb{S}_T^*(\alpha,\gamma_h^*(\alpha)) &\leq \mathbb{S}_T^*(\alpha,\gamma_{g,h}^*) = \mathbb{S}_T^*(\alpha_{g,h}^*,\gamma_{g,h}^*) + \mathbb{S}_T^*(\alpha,\gamma_{g,h}^*) - \mathbb{S}_T^*(\alpha_{g,h}^*,\gamma_{g,h}^*).
\end{aligned}$$

By definition, $\min_{\alpha,h(\gamma)=h(\widehat{\gamma})}\mathbb{S}_T^*(\alpha,\gamma) = \mathbb{S}_T^*(\alpha_{g,h}^*,\gamma_{g,h}^*)$ and $\min_{\alpha,\gamma}\mathbb{S}_T^*(\alpha,\gamma) = \mathbb{S}_T^*(\alpha_g^*,\gamma_g^*)$. Hence it suffices to show if $|\alpha-\alpha_g^*|_2 = o_{P^*}(T^{-1/2})$, and $|\alpha-\alpha_{g,h}^*|_2 = o_{P^*}(T^{-1/2})$,

$$\begin{aligned}
&\mathbb{S}_T^*(\alpha,\gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*,\gamma_g^*) \leq o_{P^*}(T^{-1}) \\
&\mathbb{S}_T^*(\alpha,\gamma_{g,h}^*) - \mathbb{S}_T^*(\alpha_{g,h}^*,\gamma_{g,h}^*) \leq o_{P^*}(T^{-1}).
\end{aligned}$$

By $|\alpha_g^* - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$ and the triangular inequality, $|\alpha - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$. Uniformly in γ so that $|\gamma - \widehat{\gamma}|_2 \leq CT^{-(1-2\varphi)}$,

$$
\begin{aligned}
& \mathbb{S}_T^*(\alpha,\gamma) - \mathbb{S}_T^*(\alpha_g^*,\gamma) = (\alpha_g^* - \alpha)'\frac{2}{T}\sum_t Z_t(\gamma)\eta_t\widehat{\varepsilon}_t + (\alpha_g^* - \alpha)'\frac{1}{T}\sum_t Z_t(\gamma)[Z_t(\widehat{\gamma}) - Z_t(\gamma)]'\widehat{\alpha} \\
& +(\alpha_g^* - \alpha)'\frac{1}{T}\sum_t Z_t(\gamma)Z_t(\gamma)'(\widehat{\alpha} - \alpha) + (\alpha_g^* - \alpha)'\frac{1}{T}\sum_t Z_t(\gamma)Z_t(\gamma)'(\widehat{\alpha} - \alpha_g^*) \\
= \;& o_{P^*}(T^{-1}) + o_{P^*}(T^{-1/2})\frac{1}{T}\sum_t Z_t(\gamma)\eta_t\widehat{\varepsilon}_t + o_{P^*}(T^{-1/2})\frac{1}{T}\sum_t Z_t(\gamma)[Z_t(\widehat{\gamma}) - Z_t(\gamma)]'\widehat{\alpha} \\
= \;& o_{P^*}(T^{-1}) + o_{P^*}(T^{-1/2})\frac{1}{T}\sum_t Z_t(\widehat{\gamma})\eta_t\widehat{\varepsilon}_t + o_{P^*}(T^{-1/2})\frac{1}{T}\sum_t [Z_t(\gamma) - Z_t(\widehat{\gamma})]\eta_t\widehat{\varepsilon}_t \\
& +o_{P^*}(T^{-1/2})\frac{1}{T}\sum_t Z_t(\gamma)[1\{f_t'\widehat{\gamma} > 0\} - 1\{f_t'\gamma > 0\}]'x_t'\widehat{\delta} \\
\leq \;& o_{P^*}(T^{-1}) + o_{P^*}(T^{-1/2}) \sup_{|\gamma - \widehat{\gamma}| < CT^{-(1-2\varphi)}} |\frac{1}{T}\sum_t [Z_t(\gamma) - Z_t(\widehat{\gamma})]\eta_t\widehat{\varepsilon}_t|_2 \\
& +o_{P^*}(T^{-1/2-\varphi}) \sup_{|\gamma - \gamma_0| < CT^{-(1-2\varphi)}} \left|\frac{1}{T}\sum_t |x_t|_2^2|1(\gamma_0) - 1(\gamma)| - \mathbb{E}|x_t|_2^2|1(\gamma_0) - 1(\gamma)|\right| \\
& +o_{P^*}(T^{-1/2-\varphi}) \sup_{|\gamma - \gamma_0| < CT^{-(1-2\varphi)}} \mathbb{E}|x_t|_2^2|1(\gamma_0) - 1(\gamma)| \\
\leq \;& o_{P^*}(T^{-1}) + o_{P^*}(T^{-1/2})T^{-(1-\varphi)} + o_{P^*}(T^{-1/2-\varphi})T^{-(1-2\varphi)} = o_{P^*}(T^{-1}).
\end{aligned} \tag{H.4}
$$

Applying the above to $\gamma = \gamma_g^*$, which satisfies $|\gamma - \widehat{\gamma}|_2 \leq CT^{-(1-2\varphi)}$ with bootstrap probability measure arbitrarily close to one by step 1 due to step 0, we have

$$\mathbb{S}_T^*(\alpha,\gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*,\gamma_g^*) = o_{P^*}(T^{-1}).$$

In addition, (H.4) also applies when α_g^* is replaced with $\alpha_{g,h}^*$. That is, $\mathbb{S}_T^*(\alpha,\gamma) - \mathbb{S}_T^*(\alpha_{g,h}^*,\gamma) = o_{P^*}(T^{-1})$ uniformly in $|\gamma - \widehat{\gamma}|_2 \leq CT^{-(1-2\varphi)}$. By step 0, $|\gamma_{g,h}^* - \widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$. Hence let $\gamma = \gamma_{m,h}^*$, we have

$$\mathbb{S}_T^*(\alpha,\gamma_{m,h}^*) - \mathbb{S}_T^*(\alpha_{g,h}^*,\gamma_{m,h}^*) = o_{P^*}(T^{-1}).$$

Proof of Step 4. Note that $|\alpha - \widehat{\alpha}| = O_{P^*}(T^{-1/2})$. The proof is then simply the bootstrap version of step 3 of the iterative estimator in the known factor case. Thus we just sketch the proof for $|\gamma_h^*(\alpha) - \widehat{\gamma}|_2 = O_{P^*}(T^{-(1-2\varphi)})$ for brevity.

For generic γ,

$$\begin{aligned}\mathbb{S}_T^*(\alpha,\gamma)-\mathbb{S}_T^*(\alpha,\widehat{\gamma}) &= \delta'\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\left|1_t(\gamma)-1_t(\widehat{\gamma})\right|\delta-\frac{2}{T}\sum_{t=1}^{T}\eta_t\widehat{\varepsilon}_tx_t'\left(1_t(\gamma)-1_t(\widehat{\gamma})\right)\delta\\ &\quad+(\alpha-\widehat{\alpha})'\frac{2}{T}\sum_{t=1}^{T}Z_t(\widehat{\gamma})x_t'\left(1_t(\gamma)-1_t(\widehat{\gamma})\right)\delta.\end{aligned}$$

Apply Lemma F.2 with γ_0 replaced by a generic γ_2, uniformly for γ,γ_2, and an arbitrarily small $\eta>0$,

$$\begin{aligned}\delta'\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\left|1_t(\gamma)-1_t(\gamma_2)\right|\delta &= O_p\left(\frac{1}{T^{1+\varphi}}\right)+|\delta|_2\,\eta T^{-2\varphi}\left|\gamma-\gamma_2\right|_2\\ &\quad+T^{-2\varphi}\mathbb{E}\left(d_0'x_t\right)^2\left|1_t(\gamma)-1_t(\gamma_2)\right|\\ &\geq O_p\left(\frac{1}{T}\right)+cT^{-2\varphi}\left|\gamma-\gamma_2\right|_2\\ \frac{2}{T}\sum_{t=1}^{T}\eta_t\widehat{\varepsilon}_tx_t'\left(1_t(\gamma)-1_t(\gamma_2)\right)\delta &\leq O_{P^*}\left(\frac{1}{T}\right)+\eta T^{-2\varphi}\left|\gamma-\gamma_2\right|_2\\ (\alpha-\widehat{\alpha})'\frac{2}{T}\sum_{t=1}^{T}Z_t(\widehat{\gamma})x_t'\left(1_t(\gamma)-1_t(\gamma_2)\right)\delta &= \left(O_p\left(\frac{1}{T}\right)+\eta T^{-2\varphi}\left|\gamma-\gamma_2\right|_2+T^{-\varphi}\left|\gamma-\gamma_2\right|_2\right)\\ &\quad\times O_{P^*}\left(T^{-1/2}\right).\end{aligned}$$

Combining these bounds and setting $\gamma=\gamma_h^*(\alpha)$, and $\gamma_2=\widehat{\gamma}$,

$$0\geq\mathbb{S}_T^*(\alpha,\gamma_h^*(\alpha))-\mathbb{S}_T^*(\alpha,\widehat{\gamma})\;\geq\;O_{P^*}\left(T^{-1}\right)+cT^{-2\varphi}\left|\gamma_h^*(\alpha)-\widehat{\gamma}\right|.$$

This implies $|\gamma_h^*(\alpha)-\widehat{\gamma}|\leq O_{P^*}(T^{-(1-2\varphi)})$. The same argument yields $|\gamma^*(\alpha)-\widehat{\gamma}|\leq O_{P^*}(T^{-(1-2\varphi)})$. ∎

H.2 Proof of Theorem 5.1: estimated factor case

H.2.1 The bootstrap with re-estimated factors using PCA

We now present details on our main bootstrap procedure that re-estimates factors in the bootstrap sample. This is given by Gonçalves and Perron (2019). To maintain the cross-sectional dependence among the idiosyncratic components in the bootstrap factor models, we rely on a consistent (in spectral norm) covariance matrix for e_t, given by $\widehat{\mathrm{var}}(e_t)$. Such a covariance estimator can be obtained via thresholding as in Fan, Liao, and Mincheva (2013). Let $\mathcal{W}_t^*$ be a sequence of independent $N\times 1$ multivariate standard normal vectors. As in Gonçalves and Perron (2019), generate bootstrap data

$$\mathcal{Y}_t^*=\widehat{\Lambda}\widetilde{f}_{1t}+\widehat{\mathrm{var}}(e_t)^{1/2}\mathcal{W}_t^*.$$

Then apply PCA to estimate factors, obtaining $\widetilde{F}_{1t}^*$ as the estimated factors in the bootstrap sample. Let $F_t^* = (F_{1t}^*, -1)'$. It has the following expansion

$$F_t^* = H_T^{*'}(\widetilde{f}_t + \frac{1}{\sqrt{N}} h_t^*) + r_t^*$$

where r_t^* is some remainder term. In addition, let V^* be the diagonal matrix containing the top $\dim(\widetilde{f}_{1t})$ eigenvalues of $\mathcal{Y}^*\mathcal{Y}^{*'}/(NT)$. Then

$$H_T^{*'} = \begin{pmatrix} H^{*'} & 0 \\ 0 & 1 \end{pmatrix}, \quad H^{*'} = V^{*-1}\frac{1}{T}\sum_t F_{1t}^* \widetilde{f}_{1t}' \widehat{S}_\Lambda, \quad h_t^* = \begin{pmatrix} h_{1t}^* \\ 0 \end{pmatrix}.$$

where $\widehat{S}_\Lambda = \frac{1}{N}\widehat{\Lambda}'\widehat{\Lambda}$, and

$$h_{1t}^* = \widehat{S}_\Lambda^{-1}\frac{1}{N}\widehat{\Lambda}\widehat{\mathrm{var}}(e_t)^{1/2}\mathcal{W}_t^* := \mathcal{Z}_t^*.$$

Note that F_t^* estimates $\widetilde{f}_t$, the "true factors" in the bootstrap sample, up to a new rotation matrix H_T^*. Hence the bootstrap distribution of F_t^* would not be able to exactly mimic the sampling distribution of $\widetilde{f}_t$. Fortunately, such a rotation discrepancy can be removed in the bootstrap estimation because H_T^* is known. As such, we can define

$$f_t^* := H_T^{*'-1} F_t^*$$

as the final "estimated factors" in the bootstrap sample, whose asymptotic bootstrap distribution would then exactly mimic that of $\tilde{f}_t$ without rotations.

H.2.2 Gaussian-perturbed factors

We now present an alternative bootstrap procedure using Gaussian-perturbed factors. Let $\{\mathcal{W}_t^* : t \leq T\}$ be a sequence of independent $K \times 1$ multivariate standard normal random vectors. We simply use

$$f_t^* := \widetilde{f}_t + N^{-1/2}\widehat{\Sigma}_h^{1/2}\mathcal{W}_t^*,$$

where $N^{-1/2}\widehat{\Sigma}_h^{1/2}\mathcal{W}_t^*$ is a Gaussian perturbation; $\widehat{\Sigma}_h$ is an estimator for the asymptotic variance of $H'h_{1t}$, as defined in (4.6)

$$\Sigma_h := \mathrm{var}\left[(\frac{1}{N}\Lambda'\Lambda)^{-1}\frac{1}{\sqrt{N}}\Lambda' e_t\right].$$

Hence we can use

$$\widehat{\Sigma}_h = N(\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}'\widehat{\mathrm{var}}(e_t)\widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Lambda})^{-1},$$

where $\widehat{\text{var}}(e_t)$ is a high-dimensional covariance estimator for $\text{var}(e_t)$. For instance, Fan, Liao, and Mincheva (2013) assumed that $\text{var}(e_t)$ is a sparse covariance matrix, and constructed $\widehat{\text{var}}(e_t)$ using thresholding.

H.2.3 Proof of the distribution of LR

Let

$$l_{NT} = \sqrt{r_{NT} T^{1+2\varphi}}.$$

Proof. Define $\widehat{\gamma}_h = \arg\min_{\alpha, h(\gamma)=0} \mathbb{S}_T(\alpha,\gamma)$, $\widehat{\alpha}(\gamma) = \arg\min_{\alpha} \mathbb{S}_T(\alpha,\gamma)$, and $\widehat{\alpha}_h = \widehat{\alpha}(\widehat{\gamma}_h)$. Then

$$\begin{aligned} l_{NT} \min_{\alpha,\gamma} \mathbb{S}_T(\alpha,\gamma) LR &= l_{NT}[\mathbb{S}_T(\widehat{\alpha}_h,\widehat{\gamma}_h) - \mathbb{S}_T(\widehat{\alpha},\widehat{\gamma})] \\ &= A_1 + A_2 - A_3 \quad \text{where,} \\ A_1 &= l_{NT}[\mathbb{S}_T(\widehat{\alpha}_h,\widehat{\gamma}_h) - \mathbb{S}_T(\widehat{\alpha}_h,\gamma_0)] \\ A_2 &= l_{NT}[\mathbb{S}_T(\widehat{\alpha}_h,\gamma_0) - \mathbb{S}_T(\widehat{\alpha},\gamma_0)] \\ A_3 &= l_{NT}[\mathbb{S}_T(\widehat{\alpha},\widehat{\gamma}) - \mathbb{S}_T(\widehat{\alpha},\gamma_0)]. \end{aligned}$$

We first analyze $\mathbb{S}_T(\alpha,\gamma) - \mathbb{S}_T(\alpha,\gamma_0)$.

We have $\mathbb{S}_T(\alpha,\gamma) - \mathbb{S}_T(\alpha,\gamma_0) = \widetilde{\mathbb{C}}_3(\alpha,\gamma) - \widetilde{\mathbb{C}}_1(\alpha,\gamma) + \sum_{d=2}^{6} \widetilde{R}_d(\alpha,\gamma)$ where $\widetilde{\mathbb{C}}_1(\alpha,\gamma)$, $\widetilde{\mathbb{C}}_3(\alpha,\gamma), \widetilde{R}_2(\alpha,\gamma), \widetilde{R}_3(\alpha,\gamma)$ are as defined in Section G.3.2, and

$$\begin{aligned} \widetilde{R}_4(\alpha,\gamma) &= \frac{1}{T}\sum_t 2\widetilde{Z}_t(\gamma_0)'(\alpha-\alpha_0)x_t'(\delta-\delta_0)(1\{\widetilde{f}_t'\gamma>0\} - 1\{\widetilde{f}_t'\gamma_0>0\}) \\ \widetilde{R}_5(\alpha,\gamma) &= \frac{1}{T}\sum_t (x_t'(\delta-\delta_0))^2 |1\{\widetilde{f}_t'\gamma>0\} - 1\{\widetilde{f}_t'\gamma_0>0\}| \\ \widetilde{R}_6(\alpha,\gamma) &= \frac{1}{T}\sum_t x_t'(\delta-\delta_0)x_t'\delta |1\{\widetilde{f}_t'\gamma>0\} - 1\{\widetilde{f}_t'\gamma_0>0\}|. \end{aligned}$$

Uniformly for $|\alpha-\alpha_0| = O_P(T^{-1/2})$ and $|\gamma-\gamma_0|_2 = O_P(r_{NT}^{-1})$, we have $l_{NT}|\widetilde{R}_d(\alpha,\gamma)| = o_P(1)$ for $d = 3 \sim 6$. In addition, $\widetilde{R}_2(\alpha,\gamma) + \widetilde{\mathbb{C}}_3(\alpha,\gamma) = \mathbb{G}_1(H_T\gamma) + o_P(l_{NT}^{-1})$ for $\mathbb{G}_1(\phi)$ as defined in (G.11), by Lemmas G.1, G.2, G.4. So

$$l_{NT}[\mathbb{S}_T(\alpha,\gamma) - \mathbb{S}_T(\alpha,\gamma_0)] = l_{NT}[\mathbb{G}_1(H_T\gamma) - \widehat{\mathbb{C}}_1(\alpha_0,\gamma)] + o_P(1).$$

Next, $\mathbb{S}_T(\widehat{\alpha}_h,\widehat{\gamma}_h) \leq \mathbb{S}_T(\alpha_0,\gamma_0)$ under $h(\gamma_0) = h(\widehat{\gamma}_h) = 0$. Thus the same proof as the rate of convergence for $(\widehat{\alpha},\widehat{\gamma})$ also carries over to prove that $|\widehat{\gamma}_h - \gamma_0|_2 = O_P(r_{NT}^{-1})$ and $|\widehat{\alpha}_h - \alpha_0|_2 = O_P(T^{-1/2})$. Now let $\alpha = \alpha_0 + a \cdot T^{-1/2}$, $\gamma = \gamma_0 + g \cdot r_{NT}^{-1}$,

$$\mathbb{K}_T(a,g) = l_{NT}\left(\mathbb{S}_T\left(\alpha_0 + a\cdot T^{-1/2}, \gamma_0 + g\cdot r_{NT}^{-1}\right) - \mathbb{S}_T(\alpha_0,\gamma_0)\right)$$

$$\begin{aligned}
&= l_{NT}\sum_{d=1}^{3}\widetilde{R}_d(\alpha,\gamma) - l_{NT}\sum_{d=1}^{2}\widetilde{\mathbb{C}}_d(\alpha,\gamma) + l_{NT}\sum_{d=3}^{4}\widetilde{\mathbb{C}}_d(\alpha,\gamma)\\
&= l_{NT}[\mathbb{G}_1(H_T\gamma) - \widehat{\mathbb{C}}_1(\alpha_0,\gamma)] + l_{NT}[\widehat{R}_1(\alpha,\gamma_0) - \widehat{\mathbb{C}}_2(\alpha,\gamma)] + o_P(1)\\
&= \mathbb{K}_{4T}(g) + \mathbb{K}_{1T}(a) + o_P(1),
\end{aligned}$$

where $\mathbb{K}_{1T}(a) := l_{NT}[\widehat{R}_1(\alpha,\gamma_0) - \widehat{\mathbb{C}}_2(\alpha,\gamma)]$, which does not depend on γ, and

$$\mathbb{K}_{4T}(g) := l_{NT}[\mathbb{G}_1(H_T(\gamma_0 + g\cdot r_{NT}^{-1})) - \widehat{\mathbb{C}}_1(\alpha_0,\gamma_0 + g\cdot r_{NT}^{-1})].$$

Define

$$\begin{aligned}
(\widehat{a}_h,\widehat{g}_h) &= \arg\min_{a,h(\gamma_0+g_h r_{NT}^{-1})=0}\mathbb{K}_T(a,g_h), \quad (\widehat{a},\widehat{g}) = \arg\min_{a,g}\mathbb{K}_T(a,g)\\
\widehat{g}_h &= r_{NT}^{-1}(\widehat{\gamma}_h - \gamma_0), \qquad \widehat{g} = r_{NT}^{-1}(\widehat{\gamma} - \gamma_0)\\
\widetilde{g}_h &= \arg\min_{h(\gamma_0+g_h r_T^{-1})=0}\mathbb{K}_{4T}(g), \qquad \widetilde{g} = \arg\min_{g}\mathbb{K}_{4T}(g).
\end{aligned}$$

Then the same proof as in Section H.1.1 shows that

$$\mathbb{K}_{4T}(\widehat{g}_h) = \mathbb{K}_{4T}(\widetilde{g}_h) + o_P(1), \quad \mathbb{K}_{4T}(\widehat{g}) = \mathbb{K}_{4T}(\widetilde{g}) + o_P(1).$$

Thus

$$\begin{aligned}
A_1 &= l_{NT}[\mathbb{G}_1(H_T\widehat{\gamma}_h) - \widehat{\mathbb{C}}_1(\alpha_0,\widehat{\gamma}_h)] + o_P(1)\\
&= \mathbb{K}_{4T}(\widehat{g}_h) + o_P(1) = \arg\min_{h(\gamma_0+g_h r_T^{-1})=0}\mathbb{K}_{4T}(g_h) + o_P(1)\\
&= \arg\min_{g_h'\nabla h=0}\mathbb{K}_{4T}(g_h) + o_P(1)\\
A_3 &= l_{NT}[\mathbb{G}_1(H_T\widehat{\gamma}) - \widehat{\mathbb{C}}_1(\alpha_0,\widehat{\gamma})] + o_P(1)\\
&= \arg\min_{g}\mathbb{K}_{4T}(g) + o_P(1).
\end{aligned}$$

Also $A_2 = o_P(1)$ following a similar proof as in Lemma H.1. Sections G.7.1 G.7.2 show that $\mathbb{K}_{4T}(\cdot) \Rightarrow \mathbb{Q}(\omega,\cdot)$, where $l_{NT}\mathbb{G}_1(H_T(\gamma_0 + g\cdot r_{NT}^{-1}))$ is the bias part and $l_{NT}\widehat{\mathbb{C}}_1(\alpha_0,\gamma_0 + g\cdot r_{NT}^{-1})$ is the empirical process part. Hence by the continuous mapping theorem,

$$l_{NT}\cdot LR \to^d \sigma_\varepsilon^{-2}\min_{g_h'\nabla h=0}\mathbb{Q}(\omega,g_h) - \sigma_\varepsilon^{-2}\min_{g}\mathbb{Q}(\omega,g).$$

∎

H.2.4 Proof of the distribution of LR_k^*

Proof. The proof below holds for both cases of bootstrap estimated factors:

re-estimated factors using PCA Recall that $f_t^* = H_T^{*'-1} F_t^*$. Then we have

$$f_t^* = \widehat{f}_t^* + H_T^{*'-1} r_t^* + m_t, \quad \widehat{f}_t^* := \widehat{f}_t + \frac{1}{\sqrt{N}} h_t^* \tag{H.5}$$

where $m_t = \widetilde{f}_t - \widehat{f}_t$, whose probability bound is given in Section G.3.1. Note that the bootstrap probability bound for r_t^* is similar to that of m_t. Also,

$$h_t^* = (h_{1t}^*, 0), \quad h_{1t}^* \sim \mathcal{N}(0, \widehat{\Sigma}_h)$$

perturbed bootstrap factors

In the case of directly using perturbed factors for bootstrap: $f_t^* = \widetilde{f}_t + N^{-1/2} h_t^*$, we can still write

$$f_t^* = \widehat{f}_t^* + H_T^{*'-1} r_t^* + m_t, \quad \widehat{f}_t^* := \widehat{f}_t + \frac{1}{\sqrt{N}} h_t^* \tag{H.6}$$

where $r_t^* = 0$ and

$$h_t^* = (h_{1t}^*, 0), \quad h_{1t}^* = \widehat{\Sigma}_h^{1/2} \mathcal{W}_t^* \sim \mathcal{N}(0, \widehat{\Sigma}_h).$$

Step 1. Expansion of $l_{NT}(\mathbb{S}_T^*(\alpha,\gamma) - \mathbb{S}_T^*(\alpha,\widehat{\gamma}))$.

We have $l_{NT}\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) LR_k^* = l_{NT}[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) - \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*)] = A_1^* + A_2^* - A_3^*$, where

$$A_1^* = l_{NT}[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) - \mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma})], A_2^* = l_{NT}[\mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}) - \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma})], A_3^* = l_{NT}[\mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) - \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma})].$$

Define

$$\begin{aligned}
\widetilde{R}_1^*(\alpha,\gamma) &= \frac{1}{T}\sum_{t=1}^T \left(Z_t^*(\gamma)'(\alpha - \widehat{\alpha})\right)^2 \\
\widetilde{R}_2^*(\gamma) &= \frac{1}{T}\sum_{t=1}^T \left(x_t'\delta_0\right)^2 |1\{f_t^{*\prime}\gamma > 0\} - 1\{f_t^{*\prime}\widehat{\gamma} > 0\}|, \\
\widetilde{R}_3^*(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^T x_t'\widehat{\delta}\left(1\left\{f_t^{*\prime}\gamma > 0\right\} - 1\left\{f_t^{*\prime}\widehat{\gamma} > 0\right\}\right) Z_t^*(\gamma)'(\alpha - \widehat{\alpha}), \\
\widetilde{R}_4^*(\alpha,\gamma) &= \frac{1}{T}\sum_t 2Z_t^*(\widehat{\gamma})'(\alpha - \widehat{\alpha})x_t'(\delta - \widehat{\delta})(1\{f_t^{*\prime}\gamma > 0\} - 1\{f_t^{*\prime}\widehat{\gamma} > 0\}) \\
\widetilde{R}_5^*(\alpha,\gamma) &= \frac{1}{T}\sum_t (x_t'(\delta - \delta_0))^2 |1\{f_t^{*\prime}\gamma > 0\} - 1\{f_t^{*\prime}\widehat{\gamma} > 0\}| \\
\widetilde{R}_6^*(\alpha,\gamma) &= \frac{2}{T}\sum_t x_t'(\delta - \delta_0)x_t'\delta |1\{f_t^{*\prime}\gamma > 0\} - 1\{f_t^{*\prime}\widehat{\gamma} > 0\}| \\
\widetilde{\mathbb{C}}_1^*(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^T \eta_t\widehat{\varepsilon}_t x_t'\delta\left(1\{f_t^{*\prime}\gamma > 0\} - 1\{f_t^{*\prime}\widehat{\gamma} > 0\}\right), \\
\widetilde{\mathbb{C}}_2^*(\alpha) &= \frac{2}{T}\sum_{t=1}^T \eta_t\widehat{\varepsilon}_t Z_t^*(\widehat{\gamma})'(\alpha - \widehat{\alpha}),
\end{aligned}$$

$$
\begin{aligned}
\widetilde{\mathbb{C}}_3^*(\alpha,\gamma) &= \frac{2}{T}\sum_{t=1}^T x_t'\widehat{\delta}x_t'\delta\left(1\{f_t^{*\prime}\widehat{\gamma}>0\}-1\{\widetilde{f}_t'\widehat{\gamma}>0\}\right)\left(1\{f_t^{*\prime}\gamma>0\}-1\{f_t^{*\prime}\widehat{\gamma}>0\}\right),\\
\widetilde{\mathbb{C}}_4^*(\alpha) &= \frac{2}{T}\sum_{t=1}^T x_t'\widehat{\delta}\left(1\{f_t^{*\prime}\widehat{\gamma}>0\}-1\{\widetilde{f}_t'\widehat{\gamma}>0\}\right)Z_t^*(\widehat{\gamma})'(\alpha-\widehat{\alpha}),\\
\widehat{\mathbb{C}}_1^*(\gamma) &= \frac{2}{T}\sum_{t=1}^T \eta_t\widehat{\varepsilon}_t x_t'\widehat{\delta}\left(1\{\widehat{f}_t^{*\prime}\gamma>0\}-1\{\widehat{f}_t^{*\prime}\widehat{\gamma}>0\}\right),\\
\widehat{R}_2^*(\gamma,\widehat{\gamma}) &= \frac{1}{T}\sum_{t=1}^T \left(x_t'\delta_0\right)^2|1\{\widehat{f}_t^{*\prime}\gamma>0\}-1\{\widehat{f}_t^{*\prime}\widehat{\gamma}>0\}|,\\
\widehat{\mathbb{C}}_3^*(\gamma,\widehat{\gamma}) &= \frac{2}{T}\sum_{t=1}^T (x_t'\delta_0)^2\left(1\{\widehat{f}_t^{*\prime}\widehat{\gamma}>0\}-1\{\widehat{f}_t'\widehat{\gamma}>0\}\right)\left(1\{\widehat{f}_t^{*\prime}\gamma>0\}-1\{\widehat{f}_t^{*\prime}\widehat{\gamma}>0\}\right).
\end{aligned}
$$

Uniformly in $|\alpha-\widehat{\alpha}|_2 = O_P(T^{-1/2})$ and $|\gamma-\widehat{\gamma}|_2 = O_P(r_{NT}^{-1})$, we have $l_{NT}|\widetilde{R}_d^*(\alpha,\gamma)| = o_{P^*}(1)$ for $d = 3\sim 6$, $l_{NT}|\widehat{\mathbb{C}}_3^*(\gamma,\widehat{\gamma}) - \widetilde{\mathbb{C}}_3^*(\alpha,\gamma)| + l_{NT}|\widehat{R}_2^*(\gamma,\widehat{\gamma}) - \widetilde{R}_2^*(\gamma)| = o_{P^*}(1)$, and $l_{NT}|\widehat{\mathbb{C}}_1^*(\gamma) - \widetilde{\mathbb{C}}_1^*(\alpha,\gamma)| = o_{P^*}(1)$. These convergences are straightforward to verify as in the original sampling space. We verify $l_{NT}|\widehat{\mathbb{C}}_1^*(\gamma) - \widetilde{\mathbb{C}}_1^*(\alpha,\gamma)| = o_{P^*}(1)$ at the end of the proof (step 5) for illustration.

Next, write

$$
\breve{g}_{t,H,\Sigma} := \breve{g}_t + N^{-1/2}H'^{-1}\Sigma^{1/2}\mathcal{W}_t^*,
$$

where $\mathcal{W}_t^*$ is standard normal. Also write

$$
\begin{aligned}
\mathbb{G}_{H,\Sigma}(\phi_2,\phi_1) \;:=\; & 2\mathbb{E}(x_t'\delta_0)^2\left(1\{\breve{g}_{t,H,\Sigma}'\phi_1>0\}-1\{\breve{g}_t'\phi_1>0\}\right)\left(1\{\breve{g}_{t,H,\Sigma}'\phi_2>0\}-1\{\breve{g}_{t,H,\Sigma}'\phi_1>0\}\right)\\
&+\mathbb{E}\left(x_t'\delta_0\right)^2|1\{\breve{g}_{t,H,\Sigma}'\phi_2>0\}-1\{\breve{g}_{t,H,\Sigma}'\phi_1>0\}|.
\end{aligned}
$$

where $\mathbb{E}$ is with respect to the joint distribution of the sampling distribution and $\mathcal{W}_t^*$. Then $\widehat{f}_t^* = H_T'\breve{g}_{t,H_T,\widehat{\Sigma}}$, and $\mathbb{G}_{H_T,\widehat{\Sigma}}(H_T\gamma, H_T\widehat{\gamma}) = \mathbb{E}(\widehat{\mathbb{C}}_3^*(\gamma_1,\gamma_2) + \widehat{R}_2^*(\gamma_1,\gamma_2))$. Then for $\phi = H_T\gamma$ and $\widehat{\phi} = H_T\widehat{\gamma}$,

$$
\begin{aligned}
l_{NT}(\mathbb{S}_T^*(\alpha,\gamma)-\mathbb{S}_T^*(\alpha,\widehat{\gamma})) &= l_{NT}(\widetilde{\mathbb{C}}_3^*(\alpha,\gamma)-\widetilde{\mathbb{C}}_1^*(\alpha,\gamma)+\sum_{d=3}^6\widetilde{R}_d^*(\alpha,\gamma)+\widetilde{R}_2^*(\gamma))\\
&= l_{NT}(\widehat{\mathbb{C}}_3^*(\gamma,\widehat{\gamma})+\widehat{R}_2^*(\gamma,\widehat{\gamma}))-l_{NT}\widehat{\mathbb{C}}_1^*(\gamma)+o_{P^*}(1)\\
&= l_{NT}\mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi})-l_{NT}\widehat{\mathbb{C}}_1^*(\gamma)+o_{P^*}(1). \qquad \text{(H.7)}
\end{aligned}
$$

Step 2. Probability limit of $l_{NT}\mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi})$.

Fix ϕ_1 in a neighborhood of ϕ_0. We first obtain a similar expansion as in (G.18).

$$
\begin{aligned}
A_{1t}^*(\phi_2,\phi_1) &= 1\left\{\breve{g}_{t,H,\Sigma}'\phi_2\le 0<\breve{g}_{t,H,\Sigma}'\phi_1\right\}1\left\{\breve{g}_t'\phi_1>0\right\}\\
A_{2t}^*(\phi_2,\phi_1) &= 1\left\{\breve{g}_{t,H,\Sigma}'\phi_1\le 0<\breve{g}_{t,H,\Sigma}'\phi_2\right\}1\left\{\breve{g}_t'\phi_1\le 0\right\}
\end{aligned}
$$

$$
\begin{aligned}
A_{3t}^*(\phi_2,\phi_1) &= 1\left\{\breve{g}_{t,H,\Sigma}'\phi_2 \le 0 < \breve{g}_{t,H,\Sigma}'\phi_1\right\}1\left\{\breve{g}_t'\phi_1 \le 0\right\} \\
A_{4t}^*(\phi_2,\phi_1) &= 1\left\{\breve{g}_{t,H,\Sigma}'\phi_1 \le 0 < \breve{g}_{t,H,\Sigma}'\phi_2\right\}1\left\{\breve{g}_t'\phi_1 > 0\right\}.
\end{aligned}
$$

Therefore, $\mathbb{G}_{H,\Sigma}(\phi_2,\phi_1) = \mathbb{E}\left(x_t'\delta_0\right)^2\left(A_{1t}^*\left(\phi_2,\phi_1\right) + A_{2t}^*\left(\phi_2,\phi_1\right) - A_{3t}^*\left(\phi_2,\phi_1\right) - A_{4t}^*\left(\phi_2,\phi_1\right)\right).$ Let us calculate A_{1t}^* first. For notational simplicity, write

$$
h_{t,H,\Sigma}^{*'} = \mathcal{W}_t^{*'}\Sigma^{1/2}H^{-1}, \quad u_{Nt} := \breve{g}_t'\phi_1.
$$

Then

$$
\begin{aligned}
\mathbb{E}\left(x_t'\delta_0\right)^2 A_{1t}^* &= \mathbb{E}\left(x_t'\delta_0\right)^2 1\left\{-h_{t,H,\Sigma}^{*'}\phi_1 < \sqrt{N}u_{Nt} \le -\sqrt{N}\breve{g}_t'\left(\phi_2-\phi_1\right) - h_{t,H,\Sigma}^{*'}\phi_1\right\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 \le 0\right\} \\
&+\mathbb{E}\left(x_t'\delta_0\right)^2 1\left\{0 < \sqrt{N}u_{Nt} \le -\sqrt{N}\breve{g}_t'\left(\phi_2-\phi_1\right) - h_{t,H,\Sigma}^{*'}\phi_1\right\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 > 0\right\} + A_{11},
\end{aligned}
$$

where the same proof as of Lemma G.6 implies $A_{11} \le \frac{CL}{NT^{2\varphi}}$, given the assumption that $p_{\breve{g}_t'\phi|h_t^*}(\cdot)$ is bounded. Let $p_{u_{Nt}|\bigstar}(\cdot) := p_{u_{Nt}|h_{t,H,\Sigma}^{*'}\phi_1, f_{2t}, x_t}(\cdot)$ denote the conditional density of u_{Nt}. Change variable $a = \sqrt{N}u$, we have,

$$
\begin{aligned}
&\mathbb{E}\left(x_t'\delta_0\right)^2 A_1 - A_{11} \\
&= \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2\int 1\left\{-h_{t,H,\Sigma}^{*'}\phi_1 < a \le -\sqrt{N}\breve{g}_t'\left(\phi_2-\phi_1\right) - h_{t,H,\Sigma}^{*'}\phi_1\right\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 \le 0\right\}p_{u_{Nt}|\bigstar}\left(\frac{a}{\sqrt{N}}\right)da \\
&+ \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2\int 1\left\{0 < a \le -\sqrt{N}\breve{g}_t'\left(\phi_2-\phi_1\right) - h_{t,H,\Sigma}^{*'}\phi_1\right\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 > 0\right\}p_{u_{Nt}|\bigstar}\left(\frac{a}{\sqrt{N}}\right)da \\
&= -\mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_{Nt}|\bigstar}(0)\breve{g}_t'\left(\phi_2-\phi_1\right)1\{\breve{g}_t'\left(\phi_2-\phi_1\right) \le 0\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 \le 0\right\} \\
&- \mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_{Nt}|\bigstar}(0)\left(\breve{g}_t'\left(\phi_2-\phi_1\right) + \frac{h_{t,H,\Sigma}^{*'}\phi_1}{\sqrt{N}}\right)1\left\{\breve{g}_t'\left(\phi_2-\phi_1\right) + \frac{h_{t,H,\Sigma}^{*'}\phi_1}{\sqrt{N}} < 0\right\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 > 0\right\} \\
&+ B_1, \qquad \text{(H.8)}
\end{aligned}
$$

where the same proof as of Lemma G.6 implies, due to $p_{u_{Nt}|\bigstar}(.)$ is Lipschitz,

$$
|B_1| \le \frac{C'}{N}T^{-2\varphi}.
$$

So the same proof as of Lemma G.6 carries over to $A_{1t}^*...A_{4t}^*$, showing that a similar expansion as in (G.18) holds: for $\Xi(a,b)$ is as defined in (G.27), for $\breve{g}_t'(\phi_2-\phi_1) = a$, $\frac{h_t^{*\prime}\phi_1}{\sqrt{N}} = b$, and $\phi_2 = \phi_1 + Hgr_{NT}^{-1}$,

$$
\begin{aligned}
&l_{NT}\mathbb{G}_{H,\Sigma}(\phi_2,\phi_1) = \mathbb{E}[(x_t'\delta_0)^2 p_{u_{Nt}|\bigstar}(0)\Xi(a,b)] + o(l_{NT}^{-1}) \\
&= l_{NT}\mathbb{E}_{|u_{Nt}=0}(x_t'\delta_0)^2 p_{u_{Nt}}(0)\left[\left|\breve{g}_t'\left(\phi_2-\phi_1\right) + \frac{h_t^{*\prime}\phi_1}{\sqrt{N}}\right| - \left|\frac{h_t^{*\prime}\phi_1}{\sqrt{N}}\right|\right] + o(1). \qquad \text{(H.9)}
\end{aligned}
$$

When $\omega \in (0, \infty]$, $l_{NT}\mathbb{G}_{H,\Sigma}(\phi_2, \phi_1) = \breve{\mathbb{C}}_{N,H,\Sigma,\phi_1}(Hg) + o(1)$, where

$$\begin{aligned} \breve{\mathbb{C}}_{N,H,\Sigma,\phi_1}(\mathfrak{g}) &:= M_\omega \mathbb{E}_{|u_{Nt}=0}(x_t'd_0)^2 p_{u_{Nt}}(0)\left(\left|\breve{g}_t'\mathfrak{g} + \zeta_\omega^{-1}h_t^{*\prime}\phi_1\right| - \left|\zeta_\omega^{-1}h_t^{*\prime}\phi_1\right|\right) \\ \breve{\mathbb{C}}_{H,\Sigma,\phi_1}(\mathfrak{g}) &:= M_\omega \mathbb{E}_{|g_t'\phi_1=0}(x_t'd_0)^2 p_{g_t'\phi_1}(0)\left(\left|g_t'\mathfrak{g} + \zeta_\omega^{-1}h_t^{*\prime}\phi_1\right| - \left|\zeta_\omega^{-1}h_t^{*\prime}\phi_1\right|\right). \end{aligned} \quad \text{(H.10)}$$

Note that $\breve{\mathbb{C}}_{N,H,\Sigma,\phi_1}(\mathfrak{g}) = M_\omega \mathbb{E}(x_t'd_0)^2 p_{u_{Nt}|x_t,g_t}(0)\left(\left|\breve{g}_t'\mathfrak{g} + \zeta_\omega^{-1}h_t^{*\prime}\phi_1\right| - \left|\zeta_\omega^{-1}h_t^{*\prime}\phi_1\right|\right)$, so by the assumption $\sup_{x_t,h_t,g_t,\phi_1}|p_{\breve{g}_t'\phi_1|x_t,f_{2t}}(0) - p_{g_t'\phi_1|x_t,f_{2t}}(0)| = o(1)$, uniformly in ϕ_1,

$$l_{NT}\mathbb{G}_{H,\Sigma}(\phi_2, \phi_1) = \breve{\mathbb{C}}_{H,\Sigma,\phi_1}(Hg) + o(1).$$

Let $\phi = \widehat{\phi} + H_T g r_{NT}^{-1}$. Note that $\widehat{\Sigma} \to^P H'\Sigma H$. By the assumption that $\breve{\mathbb{C}}_{H,\Sigma,\phi_1}(\mathfrak{g})$ is continuous in $(H, \Sigma, \phi_1, \mathfrak{g})$,

$$\begin{aligned} & l_{NT}\mathbb{G}_{H_T,\widehat{\Sigma}}(\phi, \widehat{\phi}) = \breve{\mathbb{C}}_{H_T,\widehat{\Sigma},\widehat{\phi}}(H_T g) + o_P(1) = \breve{\mathbb{C}}_{H,H'\Sigma H,\phi_0}(Hg) + o_P(1) \\ =\ & M_\omega \mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\left(\left|g_t'Hg + \zeta_\omega^{-1}\mathcal{W}_t^{*'}(H'\Sigma H)^{1/2}H^{-1}\phi_0\right| - \left|\zeta_\omega^{-1}\mathcal{W}_t^{*'}(H'\Sigma H)^{1/2}H^{-1}\phi_0\right|\right) \\ =\ & M_\omega \mathbb{E}_{|u_t=0}(x_t'd_0)^2 p_{u_t}(0)\left(\left|g_t'Hg + \zeta_\omega^{-1}\mathcal{Z}_t\right| - \left|\zeta_\omega^{-1}\mathcal{Z}_t\right|\right) \\ =\ & A(\omega, g), \end{aligned}$$

where we note that $\mathcal{W}_t^{*'}(H'\Sigma H)^{1/2}H^{-1}\phi_0 \sim \mathcal{N}(0, \sigma_h^2)$, with $\sigma_h^2 = \phi_0'\Sigma\phi_0 = \lim \text{var}(h_t'\phi_0) = \sigma_{h,x_t,g_t}^2$ in the homoskedastic case. So $\mathcal{W}_t^{*'}(H'\Sigma H)^{1/2}H^{-1}\phi_0 =^d \mathcal{Z}_t$.

When $\omega \in (0, \infty]$, we now work with (H.9), where we treat terms a_2, a_5 in the definition of $\Xi(a, b)$ as we did for (G.19) (G.20) . Here

$$\begin{aligned} a_2 &= -\breve{g}_t'(\phi_2 - \phi_1)1\{\breve{g}_t'(\phi_2 - \phi_1) \le 0\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 \le 0\right\} \\ a_5 &= \breve{g}_t'(\phi_2 - \phi_1)1\left\{\breve{g}_t'(\phi_2 - \phi_1) > 0\right\}1\{h_{t,H,\Sigma}^{*'}\phi_1 > 0\} \\ a_2' &= -\breve{g}_t'(\phi_2 - \phi_1)1\{\breve{g}_t'(\phi_2 - \phi_1) \le 0\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 > 0\right\} \\ a_5' &= \breve{g}_t'(\phi_2 - \phi_1)1\left\{\breve{g}_t'(\phi_2 - \phi_1) > 0\right\}1\{h_{t,H,\Sigma}^{*'}\phi_1 \le 0\}. \end{aligned} \quad \text{(H.11)}$$

We note that $h_{t,H,\Sigma}^{*'}\phi_1$ is symmetric around zero, due to the Gaussianity of h_t^*. Hence

$$\mathbb{P}(h_{t,H,\Sigma}^{*'}\phi_1 \le 0|x_t, \breve{g}_t) = \mathbb{P}(h_{t,H,\Sigma}^{*'}\phi_1 > 0|x_t, \breve{g}_t) = 1/2.$$

So $\mathbb{E}(x_t'\delta_0)^2 p_{u_{NT}|\bigstar}(0)a_d = \mathbb{E}(x_t'\delta_0)^2 p_{u_{NT}|\bigstar}(0)a_d'$ for $d = 2, 5$, and we reach an expansion similar to (G.21): for $\phi_2 = \phi_1 + Hgr_{NT}^{-1}$,

$$\begin{aligned} & l_{NT}\mathbb{G}_{H,\Sigma}(\phi_2, \phi_1) = o_P(1) \\ & -2l_{NT}\mathbb{E}(x_t'\delta_0)^2 p_{u_{Nt}|\bigstar}(0)\left(\breve{g}_t'(\phi_2 - \phi_1) + \frac{h_{t,H,\Sigma}^{*'}\phi_1}{\sqrt{N}}\right)1\left\{\breve{g}_t'(\phi_2 - \phi_1) + \frac{h_{t,H,\Sigma}^{*'}\phi_1}{\sqrt{N}} < 0\right\}1\left\{h_{t,H,\Sigma}^{*'}\phi_1 > 0\right\} \end{aligned}$$

$$+2l_{NT}\mathbb{E}(x_t'\delta_0)^2 p_{u_{Nt}|\bigstar}(0)\left(\breve{g}_t'(\phi_2-\phi_1)+\frac{h^{*'}_{t,H,\Sigma}\phi_1}{\sqrt{N}}\right)1\left\{\breve{g}_t'(\phi_2-\phi_1)+\frac{h^{*'}_{t,H,\Sigma}\phi_1}{\sqrt{N}}>0\right\}1\left\{h^{*'}_{t,H,\Sigma}\phi_1\le 0\right\}$$
$$= \breve{\mathbb{C}}_{N,H,\Sigma,\phi_1,2}(Hg)+o(1)$$

where we used a similar change-variable as in Step II.1 in Section G.7.2:

$$\begin{aligned}
&\breve{\mathbb{C}}_{N,H,\Sigma,\phi_1,2}(\mathfrak{g})\\
:=\;& -\widetilde{M}_{NT}2p_{u_{Nt}}(0)\mathbb{E}[(x_t'd_0)^2F_1(\breve{g}_t,x_t,\mathfrak{g})|u_{Nt}=0]+\widetilde{M}_{NT}2p_{u_{Nt}}(0)\mathbb{E}[(x_t'd_0)^2F_2(\breve{g}_t,x_t,\mathfrak{g})|u_{Nt}=0]\\
&F_1(\breve{g}_t,x_t,\mathfrak{g}):=\int\left(\breve{g}_t'\mathfrak{g}+y\right)1\left\{\breve{g}_t'\mathfrak{g}+y<0\right\}1\{y>0\}p_{h_t^{*\prime}\phi_1}(\zeta_{NT}y)dy\\
&F_2(\breve{g}_t,x_t,\mathfrak{g}):=\int\left(\breve{g}_t'\mathfrak{g}+y\right)1\left\{\breve{g}_t'\mathfrak{g}+y>0\right\}1\{y\le 0\}p_{h_t^{*\prime}\phi_1}(\zeta_{NT}y)dy.
\end{aligned}$$

Note $\zeta_{NT}\to 0$, $\widetilde{M}_{NT}\to 1$, $|p_{h_t^{*\prime}\phi_1}(\zeta_{NT}y)-p_{h_t^{*\prime}\phi_1}(0)|\le C\zeta_{NT}y$ (Gaussian densities with bounded variance), so

$$\begin{aligned}
&\mathbb{E}_{|u_{Nt}=0}(x_t'd_0)^2\int|\left(\breve{g}_t'\mathfrak{g}+y\right)1\left\{\breve{g}_t'\mathfrak{g}+y<0\right\}1\{y>0\}|p_{h_t^{*\prime}\phi_1}(\zeta_{NT}y)-p_{h_t^{*\prime}\phi_1}(0)|dy\\
\le\;& C\zeta_{NT}\mathbb{E}_{|u_{Nt}=0}(x_t'd_0)^2\int|\left(\breve{g}_t'\mathfrak{g}+y\right)|1\left\{\breve{g}_t'\mathfrak{g}+y<0\right\}1\{y>0\}ydy\\
\le\;& C\zeta_{NT}\mathbb{E}_{|u_{Nt}=0}(x_t'd_0)^2(\breve{g}_t'\mathfrak{g})^3 1\{\breve{g}_t'\mathfrak{g}<0\}=o(1).
\end{aligned}$$

In addition, by the assumption that $|p_{u_{Nt},h_t^{*\prime}\phi_1|x_t,f_{2t}}(0,0)-p_{g_t'\phi_1,h_t^{*\prime}\phi_1|x_t,f_{2t}}(0,0)|=o(1)$,

$$\begin{aligned}
\breve{\mathbb{C}}_{N,H,\Sigma,\phi_1,2}(\mathfrak{g}) &= (\mathbb{E}(x_t'd_0)^2(g_t'\mathfrak{g})^2|u_{Nt}=0,h_t^{*\prime}\phi_1=0)p_{u_{Nt},h_t^{*\prime}\phi_1}(0,0)+o(1)\\
&= \breve{\mathbb{C}}_{H,\Sigma,\phi_1,2}(\mathfrak{g})+o(1)\\
\breve{\mathbb{C}}_{H,\Sigma,\phi_1,2}(\mathfrak{g}) &:= (\mathbb{E}(x_t'd_0)^2(g_t'\mathfrak{g})^2|g_t'\phi_1=0,h_t^{*\prime}\phi_1=0)p_{g_t'\phi_1,h_t^{*\prime}\phi_1}(0,0)+o(1).
\end{aligned}$$

Let $\phi=\widehat{\phi}+H_Tgr_{NT}^{-1}$. By the assumption that $\breve{\mathbb{C}}_{H,\Sigma,\phi_1,2}(\mathfrak{g})$ is continuous in $(H,\Sigma,\phi_1,\mathfrak{g})$,

$$\begin{aligned}
&l_{NT}\mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi})=\breve{\mathbb{C}}_{H_T,\widehat{\Sigma},\widehat{\phi},2}(H_Tg)+o_P(1)=\breve{\mathbb{C}}_{H,H'\Sigma H,\phi_0,2}(Hg)+o_P(1)\\
=\;&(\mathbb{E}(x_t'd_0)^2(g_t'Hg)^2|g_t'\phi_0=0,\mathcal{W}_t^{*\prime}(H'\Sigma H)^{1/2}H^{-1}\phi_0=0)p_{g_t'\phi_0,\mathcal{W}_t^{*\prime}(H'\Sigma H)^{1/2}H^{-1}\phi_0}(0,0)+o_P(1)\\
=\;&(\mathbb{E}(x_t'd_0)^2(g_t'Hg)^2|u_t=0,\mathcal{Z}_t=0)p_{u_t,\mathcal{Z}_t}(0,0)+o_P(1)\\
=\;&A(0,g)+o_P(1).
\end{aligned}$$

Together

$$l_{NT}\mathbb{G}_{H_T,\widehat{\Sigma}}(\widehat{\phi}+H_Tgr_{NT}^{-1},\widehat{\phi})=A(\omega,g)+o_P(1).$$

Step 3. Empirical process part.

Lemma H.4 shows that in the estimated factor case, $l_{NT}\widehat{\mathbb{C}}_1^*\left(\widehat{\gamma}+r_{NT}^{-1}g\right)\Rightarrow^* 2W(g)$, where $\widehat{f}_t^*=\widehat{f}_t+N^{-1/2}\mathcal{Z}_t^*$, and $\widehat{\mathbb{C}}_1^*(\gamma)=\frac{2}{T}\sum_{t=1}^T\eta_t\widehat{\varepsilon}_t x_t'\widehat{\delta}\left(1\{\widehat{f}_t^{*\prime}\gamma>0\}-1\{\widehat{f}_t^{*\prime}\widehat{\gamma}>0\}\right)$.

Step 4. Finish the proof.

Together, we have shown that

$$l_{NT}(\mathbb{S}_T^*(\widehat{\alpha}+aT^{-1/2},\widehat{\gamma}+r_{NT}^{-1}g)-\mathbb{S}_T^*(\widehat{\alpha}+aT^{-1/2},\widehat{\gamma})) \quad = \quad \mathbb{K}_{4T}^*(g)+o_{P^*}(1), \quad \text{(H.12)}$$

and $\mathbb{K}_{4T}^*(.) \Rightarrow^* \mathbb{Q}(\omega,.)$, where

$$\mathbb{K}_{4T}^*(g) := l_{NT}[\mathbb{G}_{H_T,\widehat{\Sigma}}(H_T(\widehat{\gamma}+r_{NT}^{-1}g),H_T\widehat{\gamma})-\widehat{\mathbb{C}}_1^*\left(\delta,\widehat{\gamma}+r_{NT}^{-1}g\right)].$$

In addition, let $\mathbb{K}_1^*(a) := l_{NT}[\widetilde{R}_1^*(\widehat{\alpha}+a\cdot T^{-1/2},\widehat{\gamma})-\widetilde{\mathbb{C}}_2^*(\widehat{\alpha}+a\cdot T^{-1/2})]$ and

$$\begin{aligned}\mathbb{K}_T^*(a,g) &:= l_{NT}\left(\mathbb{S}_T^*\left(\widehat{\alpha}+a\cdot T^{-1/2},\widehat{\gamma}+g\cdot r_{NT}^{-1}\right)-\mathbb{S}_T^*(\widehat{\alpha},\widehat{\gamma})\right)\\ &= l_{NT}\sum_{d=1}^{3}\widetilde{R}_d^*(\alpha,\gamma)-l_{NT}\sum_{d=1}^{2}\widetilde{\mathbb{C}}_d^*(\alpha,\gamma)+l_{NT}\sum_{d=3}^{4}\widetilde{\mathbb{C}}_d^*(\alpha,\gamma)\\ &= \mathbb{K}_{4T}^*(g)+\mathbb{K}_1^*(a)+o_P(1).\end{aligned}$$

By Lemma H.3, $|\widehat{\alpha}^*-\widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$, and $|\widehat{\gamma}_h^*-\widehat{\gamma}|_2 = O_P(r_{NT}^{-1})$. Define

$$\begin{aligned}\widehat{a}^* &= \sqrt{T}(\widehat{\alpha}_h^*-\widehat{\alpha}), \quad \widehat{g}_h^* = r_{NT}(\widehat{\gamma}_h^*-\widehat{\gamma})\\ \widetilde{g}_h^* &:= \arg\min_{h(\widehat{\gamma}+gr_{NT}^{-1})=h(\widehat{\gamma})}\mathbb{K}_{4T}^*(g).\end{aligned}$$

Then because $h(\widehat{\gamma}+\widetilde{g}_h^*r_{NT}^{-1})=h(\widehat{\gamma})$,

$$\mathbb{K}_{4T}^*(\widehat{g}_h^*)+\mathbb{K}_1^*(\widehat{a}^*)+o_P(1) = \mathbb{K}_T^*(\widehat{a}^*,\widehat{g}_h^*) \leq \mathbb{K}_T^*(\widehat{a}^*,\widetilde{g}_h^*)+o_{P^*}(1) = \mathbb{K}_{4T}^*(\widetilde{g}_h^*)+\mathbb{K}_1^*(\widehat{a}^*)+o_{P^*}(1).$$

where the inequality is due to Lemma H.3 that $\mathbb{S}_T^*(\widehat{\alpha}_h^*,\widehat{\gamma}_h^*) \leq \min_{\alpha,h(\gamma)=h(\widehat{\gamma})}\mathbb{S}_T^*(\alpha,\gamma)+o_{P^*}(l_{NT}^{-1})$. This implies $\mathbb{K}_{4T}^*(\widehat{g}_h^*)=\mathbb{K}_{4T}^*(\widetilde{g}_h^*)+o_{P^*}(1)$. Therefore, by (H.12),

$$\begin{aligned}A_1^* &= l_{NT}[\mathbb{S}_T^*(\widehat{\alpha}_h^*,\widehat{\gamma}_h^*)-\mathbb{S}_T^*(\widehat{\alpha}_h^*,\widehat{\gamma})] = \mathbb{K}_{4T}^*(\widehat{g}_h^*)+o_{P^*}(1),\\ &= \min_{h(\widehat{\gamma}+gr_{NT}^{-1})=h(\widehat{\gamma})}\mathbb{K}_{4T}^*(g)+o_{P^*}(1) = \min_{g_h'\nabla h=0}\mathbb{K}_{4T}^*(g)+o_{P^*}(1)\\ &\to^{d^*} \min_{g_h'\nabla h=0}\mathbb{Q}(\omega,g_h).\end{aligned}$$

As for A_2^*, it follows from $\widehat{\alpha}_h^*-\widehat{\alpha}^*=o_{P^*}(T^{-1/2})$ that $A_2^*=o_{P^*}(1)$.

Next, by Lemma H.3, $|\widehat{\gamma}^*-\widehat{\gamma}|_2=O_P(r_{NT}^{-1})$ and $\mathbb{S}_T^*(\widehat{\alpha}^*,\widehat{\gamma}^*) \leq \min_{\alpha,\gamma}\mathbb{S}_T^*(\alpha,\gamma)+o_{P^*}(l_{NT}^{-1})$, we have $A_3^* \to^{d^*} \min_g \mathbb{Q}(\omega,g)$. Hence

$$l_{NT}\mathbb{S}_T^*(\widehat{\alpha}^*,\widehat{\gamma}^*)LR^* \to^{d^*} \min_{g_h'\nabla h=0}\mathbb{Q}(\omega,g_h)-\min_g\mathbb{Q}(\omega,g).$$

This finishes the proof since $\mathbb{S}_T^*(\widehat{\alpha}^*,\widehat{\gamma}^*) \to^{P^*} \sigma^2$.

Step 5. verify $l_{NT}|\widehat{\mathbb{C}}_1^*(\gamma) - \widetilde{\mathbb{C}}_1^*(\alpha,\gamma)| = o_{P^*}(1)$.

Note that we can bound $|\widehat{\epsilon}_t| \leq |\epsilon_t| + C|x_t|_2$ with high probability. Thus for $w_t := 2|\eta_t\varepsilon_t||x_t|_2 + 2|\eta_t||x_t|_2^2$, uniformly for $|\gamma-\gamma_0|_2 < Cr_{NT}^{-1}$ and $|\alpha-\alpha_0|_2 < CT^{-1/2}$,

$$\begin{aligned} b &:= \frac{1}{T}\sum_{t=1}^T \eta_t\widehat{\varepsilon}_t x_t'\delta[1\{f_t^{*\prime}\gamma > 0\} - \widehat{f}_t^{*\prime}\gamma > 0\} \leq \frac{1}{T}\sum_{t=1}^T |\eta_t\widehat{\varepsilon}_t x_t'\delta| 1\{0 < |\widehat{f}_t^{*\prime}\gamma| < C|\widetilde{f}_t - \widehat{f}_t|_2\} \\ &\leq \frac{1}{T}\sum_{t=1}^T |\eta_t\widehat{\varepsilon}_t x_t'\delta| 1\{0 < \inf_\gamma |\widehat{f}_t^{*\prime}\gamma| < C\Delta_f\} + \mathbb{P}(|\widetilde{f}_t - \widehat{f}_t|_2 > \Delta_f)^{1/2} \\ &\leq o(l_{NT}^{-1}) + O_P(T^{-\varphi})\mathbb{P}(0 < \inf_\gamma |\widehat{f}_t^{*\prime}\gamma| < C\Delta_f) = o_P(l_{NT}^{-1}) \end{aligned}$$

given that $\inf_\gamma |\widehat{f}_t^{*\prime}\gamma|$ has a density bounded and continuous at zero. Next, write

$$a_T := \frac{2}{T}\sum_{t=1}^T \eta_t\widehat{\varepsilon}_t x_t' \left(1\{\widehat{f}_t^{*\prime}\gamma > 0\} - 1\{\widehat{f}_t^{*\prime}\widehat{\gamma} > 0\}\right).$$

Then $\mathbb{E}^* a_T = 0$, where $\mathbb{E}^*$ is the conditional expectation with respect to the distribution of $(\eta_t, \mathcal{W}_t^*)$, and note that $\eta_t, \mathcal{W}_t^*$ are independent. Now we apply Lemma I.2 to the bootstrap distribution, to reach $a_T = O_{P^*}(T^{-\varphi})\left[|\gamma-\widehat{\gamma}|_2 + \frac{1}{T^{1-2\varphi}}\right]$.

Thus $|\widehat{\mathbb{C}}_1^*(\gamma) - \widetilde{\mathbb{C}}_1^*(\alpha,\gamma)| \leq b + |a_T|_2|\delta - \widehat{\delta}|_2 = o_{P^*}(l_{NT}^{-1})$. ∎

Recall
$$r_{NT}^{-1} = \max\left(\frac{1}{(NT^{1-2\varphi})^{1/3}}, \frac{1}{T^{1-2\varphi}}\right).$$

Lemma H.3. *In the estimated factor case, the k-step bootstrap estimators $(\widehat{\alpha}^*, \widehat{\gamma}^*, \widehat{\gamma}_h^*)$ satisfy:*

$$\begin{aligned} \mathbb{S}_T^*(\widehat{\alpha}^*, \widehat{\gamma}^*) &\leq \min_{\alpha,\gamma} \mathbb{S}_T^*(\alpha,\gamma) + o_{P^*}(l_{NT}^{-1}). \\ \mathbb{S}_T^*(\widehat{\alpha}_h^*, \widehat{\gamma}_h^*) &\leq \min_{\alpha, h(\gamma)=h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha,\gamma) + o_{P^*}(l_{NT}^{-1}), \quad h(\widehat{\gamma}_h^*) = h(\widehat{\gamma}) \\ |\widehat{\alpha}^* - \widehat{\alpha}|_2 &= O_{P^*}(T^{-1/2}), \quad |\widehat{\alpha}_h^* - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2}), \quad |\widehat{\alpha}^* - \widehat{\alpha}_h^*|_2 = o_{P^*}(T^{-1/2}) \\ |\widehat{\gamma}_h^* - \widehat{\gamma}|_2 &= O_P(r_{NT}^{-1}), \quad |\widehat{\gamma}^* - \widehat{\gamma}|_2 = O_P(r_{NT}^{-1}). \end{aligned}$$

Proof. Define

$$\begin{aligned} (\alpha_g^*, \gamma_g^*) &= \arg\min \mathbb{S}_T^*(\alpha,\gamma), \quad (\alpha_{g,h}^*, \gamma_{g,h}^*) = \arg\min_{\alpha, h(\gamma)=h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha,\gamma), \\ \alpha^*(\gamma) &= \arg\min_\alpha \mathbb{S}_T^*(\alpha,\gamma), \\ \gamma^*(\alpha) &= \arg\min_\gamma \mathbb{S}_T^*(\alpha,\gamma), \quad \gamma_h^*(\alpha) = \arg\min_{\gamma: h(\gamma)=h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha,\gamma). \end{aligned}$$

Also recall the following definitions

$$\begin{aligned}\widetilde{Z}_t(\gamma) &= (x_t', x_t' 1\{\widetilde{f}_t'\gamma > 0\})' \\ \widehat{Z}_t(\gamma) &= (x_t', x_t' 1\{\widehat{f}_t'\gamma > 0\})' \\ Z_t^*(\gamma) &= (x_t', x_t' 1\{f_t^{*\prime}\gamma > 0\})' \\ \widehat{Z}_t^*(\gamma) &= (x_t', x_t' 1\{\widehat{f}_t^{*\prime}\gamma > 0\})',\end{aligned}$$

where

$$\begin{aligned}\widetilde{f}_t &= \text{estimated factors in the original data world,} \\ f_t^* &= \text{estimated factors in the bootstrap data world,} \\ \widehat{f}_t &= H_T f_t + H_T \frac{1}{\sqrt{N}} h_t, \\ \widehat{f}_t^* &= \widehat{f}_t + \frac{1}{\sqrt{N}} h_t^*.\end{aligned}$$

Our proof is divided into the following steps.

step 0: $|\gamma_{g,h}^* - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$, $|\gamma_g^* - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$ and $|\alpha_g^* - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$.

Step 0 is regarding the statistical convergence of the global minimiums in the bootstrap sample. So the proof is the same as that for $|\widehat{\gamma} - \gamma_0|_2$ and $|\widehat{\alpha} - \alpha_0|_2$.

step 1: if $|\gamma - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$, then $|\alpha^*(\gamma) - \widehat{\alpha}|_2 = O_{P^*}(T^{-1/2})$. In addition, $|\alpha^*(\gamma) - \alpha_g^*|_2 = o_{P^*}(T^{-1/2})$, and $|\alpha^*(\gamma) - \alpha_{g,h}^*|_2 = o_{P^*}(T^{-1/2})$.

In the proof of step 1 we shall show that $\alpha^*(\gamma)$ is an oracle estimator in the sense:

$$\alpha^*(\gamma) - \widehat{\alpha} = [\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t^*(\widehat{\gamma})\widehat{Z}_t^*(\widehat{\gamma})']^{-1}\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t^*(\widehat{\gamma})\eta_t\widehat{\varepsilon}_t + o_{P^*}(T^{-1/2}). \quad \text{(H.13)}$$

Write $\mathcal{A}^*(\gamma) = \frac{1}{T}\sum_{t=1}^T Z_t^*(\gamma) Z_t^*(\gamma)' = \frac{1}{T}\sum_{t=1}^T \widehat{Z}_t^*(\gamma)\widehat{Z}_t^*(\gamma)' + o_{P^*}(T^{-1/2})$. Note that $y_t^* = \widetilde{Z}_t(\widehat{\gamma})'\widehat{\alpha} + \eta_t\widehat{\varepsilon}_t$,

$$\begin{aligned}\alpha^*(\gamma) - \widehat{\alpha} &= \mathcal{A}^*(\gamma)^{-1}\frac{1}{T}\sum_{t=1}^T Z_t^*(\gamma)\, y_t^* - \widehat{\alpha} \\ &= \mathcal{A}^*(\gamma)^{-1}(a_1 + a_2), \quad \text{where} \\ a_1 &= \frac{1}{T}\sum_{t=1}^T Z_t^*(\gamma)\eta_t\widehat{\varepsilon}_t \\ a_2 &= \frac{1}{T}\sum_{t=1}^T Z_t^*(\gamma) x_t'\widehat{\delta}\left(1\left(\widetilde{f}_t'\widehat{\gamma} > 0\right) - 1\left(f_t^{*\prime}\gamma > 0\right)\right) \quad \text{(H.14)}\end{aligned}$$

By the same argument in the proof of lemma G.7,

$$a_1 = \frac{1}{T}\sum_{t=1}^{T} \widehat{Z}_t^*(\gamma)\,\eta_t\widehat{\varepsilon}_t + o_{P^*}(T^{-1/2}) = \frac{1}{T}\sum_{t=1}^{T} \widehat{Z}_t^*(\widehat{\gamma})\,\eta_t\widehat{\varepsilon}_t + o_{P^*}(T^{-1/2}).$$

On the other hand, by lemma I.2, uniformly in γ, since $T = O(N)$,

$$\begin{aligned} a_2 &= \frac{1}{T}\sum_{t=1}^{T} \widehat{Z}_t(\gamma)\,x_t'\widehat{\delta}\left(1\left(\widehat{f}_t'\widehat{\gamma} > 0\right) - 1\left(\widehat{f}_t'\gamma > 0\right)\right) + o_{P^*}(T^{-1/2}) + o_{P^*}(T^{-\varphi}N^{-1/2}) \\ &= \mathbb{E}\widehat{Z}_t(\gamma)\,x_t'\left(1\left(\widehat{f}_t'\widehat{\gamma} > 0\right) - 1\left(\widehat{f}_t'\gamma > 0\right)\right)\widehat{\delta} + \eta T^{-2\varphi}|\gamma - \widehat{\gamma}|_2 + o_{P^*}(T^{-1/2}) \\ &\leq O(T^{-\varphi})|\gamma - \widehat{\gamma}|_2 + o_{P^*}(T^{-1/2}) = o_{P^*}(T^{-1/2}). \end{aligned} \tag{H.15}$$

This gives rise to (H.13).

Now $\alpha_{g,h}^* = \alpha^*(\gamma_g^*)$ $\alpha_g^* = \alpha^*(\gamma_{g,h}^*)$ while $|\gamma_{g,h}^* - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$, $|\gamma_g^* - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$ by step 0. Hence both $\alpha_{g,h}^*$ and α_g^* satisfy the expansion (H.13), whose leading term does not depend on the general choice of γ.

step 2: if $|\alpha - \alpha_g^*|_2 = o_{P^*}(T^{-1/2})$, and $|\alpha - \alpha_{g,h}^*|_2 = o_{P^*}(T^{-1/2})$, then $\mathbb{S}_T^*(\alpha, \gamma^*(\alpha)) \leq \min_{\alpha,\gamma} \mathbb{S}_T^*(\alpha, \gamma) + o_{P^*}(l_{NT}^{-1})$, and $\mathbb{S}_T^*(\alpha, \gamma_h^*(\alpha)) \leq \min_{\alpha, h(\gamma) = h(\widehat{\gamma})} \mathbb{S}_T^*(\alpha, \gamma) + o_{P^*}(l_{NT}^{-1})$.

Note that $\mathbb{S}_T^*(\alpha, \gamma^*(\alpha)) \leq \min_{\alpha,\gamma} \mathbb{S}_T^*(\alpha, \gamma) + \mathbb{S}_T^*(\alpha, \gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*, \gamma_g^*)$. So we need to bound $\mathbb{S}_T^*(\alpha, \gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*, \gamma_g^*)$. From (H.7),

$$\begin{aligned} & \mathbb{S}_T^*(\alpha, \gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*, \gamma_g^*) \\ =\; & (\mathbb{S}_T^*(\alpha, \gamma_g^*) - \mathbb{S}_T^*(\alpha, \widehat{\gamma})) - (\mathbb{S}_T^*(\alpha_g^*, \gamma_g^*) - \mathbb{S}_T^*(\alpha_g^*, \widehat{\gamma})) + (\mathbb{S}_T^*(\alpha, \widehat{\gamma}) - \mathbb{S}_T^*(\alpha_g^*, \widehat{\gamma})) \\ =\; & \mathbb{S}_T^*(\alpha, \widehat{\gamma}) - \mathbb{S}_T^*(\alpha_g^*, \widehat{\gamma}) + o_{P^*}(l_{NT}^{-1}). \end{aligned}$$

Now given that $|\alpha - \alpha_g^*|_2 = o_{P^*}(T^{-1/2})$,

$$\begin{aligned} & \mathbb{S}_T^*(\alpha, \widehat{\gamma}) - \mathbb{S}_T^*(\alpha_g^*, \widehat{\gamma}) = \frac{1}{T}\sum_{t=1}^{T}\left(Z_t^*(\widehat{\gamma})'\left(\alpha - \alpha_g^*\right)\right)^2 + (\widehat{\alpha} - \alpha_g^*)'\frac{2}{T}\sum_{t=1}^{T}\widetilde{Z}_t(\widehat{\gamma})Z_t^*(\widehat{\gamma})(\alpha_g^* - \alpha) \\ & + \frac{2}{T}\sum_{t=1}^{T}\eta_t\widehat{\varepsilon}_t Z_t^*(\widehat{\gamma})'\left(\alpha_g^* - \alpha\right) + \frac{2}{T}\sum_{t=1}^{T}x_t'\delta_g^*\left(1\{f_t^{*\prime}\widehat{\gamma} > 0\} - 1\{\widetilde{f}_t'\widehat{\gamma} > 0\}\right)Z_t^*(\widehat{\gamma})'\left(\alpha - \alpha_g^*\right) \\ =\; & o_{P^*}(l_{NT}^{-1}), \end{aligned}$$

The same result applies when α_g^* is replaced with $\alpha_{g,h}^*$.

step 3: if $\alpha = \widehat{\alpha} + O_{P^*}\left(T^{-1/2}\right)$, then $|\gamma^*(\alpha) - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$ and $|\gamma_h^*(\alpha) - \widehat{\gamma}|_2 = O_{P^*}(r_{NT}^{-1})$.

Note that for any γ, α,

$$\mathbb{S}_T^*(\alpha,\gamma) - \mathbb{S}_T^*(\alpha,\widehat{\gamma}) = \widetilde{R}_1^*(\alpha,\gamma) - \widetilde{R}_1^*(\alpha,\widehat{\gamma}) + \widetilde{R}_2^*(\gamma) + \widetilde{R}_3^*(\alpha,\gamma) - \widetilde{\mathbb{C}}_1^*(\alpha,\gamma) + \widetilde{\mathbb{C}}_3^*(\alpha,\gamma). \tag{H.16}$$

We divide the proof of step 3 into the following sub-steps.

step 3-i: preliminary rate of convergence

We now study each term on the right of (H.16) using similar arguments of lemma G.1 and G.2. Uniformly in γ,

$$\begin{aligned}
|\widetilde{R}_1^*(\alpha,\gamma) - \widetilde{R}_1^*(\alpha,\widehat{\gamma})| &= |\alpha-\widehat{\alpha}|_2^2 O_{P^*}(\Delta_f + T^{-6} + T^{-1+\varphi}) + C|\gamma-\widehat{\gamma}|_2|\alpha-\widehat{\alpha}|_2^2 + T^{-2\varphi}O_{P^*}(\Delta_f+T^{-6}).\\
\widetilde{R}_3(\alpha,\gamma) &\le [|\alpha-\widehat{\alpha}|_2^2 + T^{-2\varphi}]O_{P^*}(\Delta_f+T^{-6}) + \left(O_{P^*}\left(T^{-1}\right) + CT^{-\varphi}\left|\gamma-\widehat{\gamma}\right|_2\right)\left|\alpha-\widehat{\alpha}\right|_2.\\
\widetilde{\mathbb{C}}_1^*(\alpha,\gamma) &= \widehat{\mathbb{C}}_1^*(\gamma) + (T^{-\varphi} + |\alpha-\widehat{\alpha}|_2)O_{P^*}(\Delta_f+T^{-6})\\
&\quad + \left(O_{P^*}\left(T^{-1}\right) + \eta T^{-2\varphi}\left|\gamma-\widehat{\gamma}\right|\right)T^{\varphi}\left|\alpha-\widehat{\alpha}\right|_2.\\
\widetilde{R}_2^*(\gamma) + \widetilde{\mathbb{C}}_3^*(\alpha,\gamma) &= \mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi}) + (T^{-\varphi}+|\alpha-\widehat{\alpha}|_2)O_{P^*}(\Delta_f+T^{-6}) + T^{-\varphi}\left|\alpha-\widehat{\alpha}\right|_2 O_{P^*}\left(N^{-1/2}\right)
\end{aligned}$$

where $\mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi})$ is as defined in (H.7).

Putting together, $\alpha = \widehat{\alpha} + O_{P^*}\left(T^{-1/2}\right)$, for $\gamma = \gamma^*(\alpha)$, $\mathbb{S}_T^*(\alpha,\gamma) - \mathbb{S}_T^*(\alpha,\widehat{\gamma}) \le 0$ implies

$$\mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi}) \le \widehat{\mathbb{C}}_1^*(\gamma) + o_{P^*}(T^{-2\varphi})N^{-1/2} + o_{P^*}(T^{-1}) + T^{-\varphi}O_{P^*}(\Delta_f+T^{-6}) + o_{P^*}(T^{-2\varphi})\left|\gamma-\widehat{\gamma}\right|_2. \tag{H.17}$$

By a similar argument as Lemmas G.4, G.5,

$$\widehat{\mathbb{C}}_1^*(\gamma) \le b_{NT}, \quad \mathbb{G}_{H_T,\widehat{\Sigma}}(\phi,\widehat{\phi}) \ge CT^{-2\varphi}|\phi-\widehat{\phi}|_2 - \frac{C}{\sqrt{N}T^{2\varphi}},$$

where for an arbitrarily small $\eta > 0$, $b_{NT} = O_{P^*}(T^{-1}) + \eta T^{-2\varphi}\left|\gamma-\widehat{\gamma}\right|_2$. Then

$$\begin{aligned}
&CT^{-2\varphi}|\phi-\widehat{\phi}|_2\\
\le\ & o_{P^*}(T^{-2\varphi})N^{-1/2} + O_{P^*}(T^{-1}) + T^{-\varphi}O_{P^*}(\Delta_f+T^{-6}) + \eta T^{-2\varphi}\left|\phi-\phi_0\right|_2 + \frac{C}{\sqrt{N}T^{2\varphi}}.
\end{aligned}$$

Since $\eta > 0$ is arbitrarily small, we have

$$|\gamma^*(\alpha) - \widehat{\gamma}|_2 = O_{P^*}(|\phi-\widehat{\phi}|_2) \le O_{P^*}(N^{-1/2} + T^{-(1-2\varphi)}).$$

step 3-ii: sharp rate of convergence

The preliminary rate is sharp when $\sqrt{N} \gg T^{1-2\varphi}$. Now suppose $\sqrt{N} = O(T^{1-2\varphi})$. By proofs similar to lemmas G.4, G.6, for $\phi = H_T\widehat{\gamma}(\alpha)$,

$$\widehat{\mathbb{C}}_1^*(\gamma) \le a_{NT}, \quad \mathbb{G}_{H_T,\widehat{\Sigma}}(\phi, \widehat{\phi}) \ge CT^{-2\varphi}\sqrt{N}|\phi - \widehat{\phi}|_2^2 - O_P(\frac{1}{T^{2\varphi}N^{5/6}}),$$

where for an arbitrarily small $\eta > 0$, $a_{NT} = T^{-2\varphi}O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right) + T^{-2\varphi}\eta\,|\phi - \phi_0|_2^2\sqrt{N}$. Substituting these bounds to (H.17) yields

$$|\gamma^*(\alpha) - \widehat{\gamma}|_2 = O_{P^*}(|\phi - \widehat{\phi}|_2) \le O_{P^*}\left(\frac{1}{(NT^{1-2\varphi})^{1/3}}\right).$$

Combining with the rates proved in claim 3, we obtain the desired result.

∎

Lemma H.4. *(i) In the known factor case, $\mathbb{K}_{3T}^*(g) \Rightarrow^* 2W(g)$, where*

$$\mathbb{K}_{3T}^*(g) := -2\sum_{t=1}^{T}\eta_t\widehat{\varepsilon}_t x_t'\widehat{\delta}\left(1_t\left(\widehat{\gamma} + g\cdot r_T^{-1}\right) - 1_t(\widehat{\gamma})\right).$$

(ii) In the estimated factor case, $\sqrt{r_{NT}T^{1+2\varphi}}\widehat{\mathbb{C}}_1^*\left(\widehat{\gamma} + r_{NT}^{-1}g\right) \Rightarrow^* 2W(g)$, where $\widehat{\mathbb{C}}_1^*(\gamma) = \frac{2}{T}\sum_{t=1}^{T}\eta_t\widehat{\varepsilon}_t x_t'\widehat{\delta}\left(1\{\widehat{f}_t^{*\prime}\gamma > 0\} - 1\{\widehat{f}_t^{*\prime}\widehat{\gamma} > 0\}\right)$, $\widehat{f}_t^* = \widehat{f}_t + N^{-1/2}\mathcal{Z}_t^*$, and $\mathcal{Z}_t^*$ is iid $\mathcal{N}(0, \widehat{\Sigma}_h)$.

Proof. (i) We first show the stochastic equicontinuity of $\mathbb{K}_{3T}^*(g)$, for which it is sufficient to show that of $\sum_{t=1}^{T}\eta_t\varepsilon_t x_t'\delta_0\left(1_t\left(\gamma_T + g\cdot r_T^{-1}\right) - 1_t(\gamma_T)\right)$ for any $\gamma_T \to \gamma_0$ since $\widehat{\delta} - \delta_0 = O_P(T^{-1/2})$, $\widehat{\gamma}$ is consistent, and $\widehat{\varepsilon}_t = \varepsilon_t + remainder_t$, where the remainder terms are treated as before. However, we can apply the maximal inequality in Lemma I.1 here since η_t is a centered iid sequence independent of the other variables. Next, to derive the finite dimensional convergence we can apply the conditional CLT e.g. Hall and Heyde (1980) for the MDS. The conditions are checked similarly as in Section F.3.

(ii) The argument for the stochastic equicontinuity is similar to the case (i). Also, the derivation in Section G.7.1 and the proof of Lemma G.9 in particular reveals that the finite dimensional limits are not affected by the change of $\widehat{f}_t$ by $\widehat{f}_t^* = \widehat{f}_t + N^{-1/2}\mathcal{Z}_t^*$.

∎

H.3 Proof of Theorem D.1

Proof of Theorem D.1. We begin with the known factor case. For each γ, our $Q_T(\gamma)$ corresponds to a modified version of the Wald statistic $T_n(\gamma)$ used in Hansen (1996). Specifically, let $\widehat{\alpha}(\gamma) = \arg\min_{\alpha} \mathbb{S}_T(\alpha, \gamma)$ and $R = (0_{d_x}, I_{d_x})$. Then it can be proved that

$$\min_{\alpha:\delta=0} \mathbb{S}_T(\alpha,\gamma) - \min_{\alpha,\gamma} \mathbb{S}_T(\alpha,\gamma) = \widehat{\alpha}(\gamma)' R' [R(\sum_t Z_t(\gamma) Z_t(\gamma)')^{-1} R']^{-1} R \widehat{\alpha}(\gamma).$$

We then replace the term $\widehat{V}_n(\gamma)$ in Hansen (1996) with

$$\widehat{V}_n(\gamma) = \frac{1}{T} \sum_{t=1}^{T} x_t x_t' 1\left\{f_t'\gamma > 0\right\} \mathbb{S}_T(\widehat{\alpha}, \widehat{\gamma}). \tag{H.18}$$

We now verify regularity conditions imposed by Hansen (1996). His Assumption 1 concerns the mixing and moment conditions that are satisfied by our Assumption 3.1 (with $v = r = 2$ in the notation used in Hansen (1996)). His Assumption 2 is a sufficient condition to ensure the tightness of the empirical process $T^{-1/2} \sum_{t=1}^{T} x_t 1\left\{f_t'\gamma > 0\right\} \varepsilon_t$, which is guaranteed by our maximal inequality Lemma I.1. Finally, his Assumption 3 follows from the ULLN. Then, the theorem is proved with the replaced $\widehat{V}_n(\gamma)$ in (H.18).

Turning to the estimated factor case, we need to establish the asymptotic equivalence between the known and unknown factors. For this purpose, it suffices to show that

$$\sup_{\gamma} \left| \frac{1}{T} \sum_{t=1}^{T} x_t x_t' \left(1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{f_t'\gamma > 0\right\}\right) \right| = o_P(1), \tag{H.19}$$

$$\sup_{\gamma} \left| \frac{1}{T} \sum_{t=1}^{T} x_t x_t' \left(1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{f_t'\gamma > 0\right\}\right) \varepsilon_t^2 \right| = o_P(1), \tag{H.20}$$

$$\sup_{\gamma} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \left(1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{f_t'\gamma > 0\right\}\right) \varepsilon_t \right| = o_P(1). \tag{H.21}$$

Recall that $\widehat{f}_t$ is defined as $\widehat{f}_t = H_T'(g_t + h_t/\sqrt{N})$. The last condition (H.21) follows directly if we show that

$$\sup_{\gamma} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \left(1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{\widehat{f}_t'\gamma > 0\right\}\right) \varepsilon_t \right| = o_P(1) \tag{H.22}$$

and

$$\sup_{\gamma} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \left(1\left\{\widehat{f}_t'\gamma > 0\right\} - 1\left\{f_t'\gamma > 0\right\}\right) \varepsilon_t \right| = o_P(1). \tag{H.23}$$

By Lemma G.1, (H.22) follows. To show (H.23), note that in view of the maximal inequality in Lemma I.1, Theorem 16.1 of Billingsley (1968) and the extended CMT in Lemma I.4, the empirical process

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t 1\left\{\widehat{f}_t'\gamma>0\right\}\varepsilon_t$$

is stochastically equicontinuous. This implies (H.23). The other two conditions (H.19) and (H.20) can be shown similarly and thus omitted. In case of the estimated factors, $f_t = H_T' g_t$ and $\gamma = H_T^{-1}\phi$ and the supremum is understood as taken over ϕ after cancelling out H_T in $f_t'\gamma$. Finally, the CMT yields the desired result. ∎

I Technical Lemmas

This section proves technical lemmas, which are repeatedly used to prove main theorems. Their proofs are given in the subsequent subsection. They are proven under the following assumption.

Assumption I.1. *Assume that $\{z_t, q_t\}_{t=1}^T$ be a sequence of strictly stationary, ergodic, and ρ-mixing array with $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$, $\mathbb{E}\left|z_t\right|_2^4<\infty$, and, for all γ in a neighborhood of γ_0, $\mathbb{E}\left(\left|z_t\right|^4|q_t=\gamma\right)<C<\infty$ and $q_t'\gamma$ has a density that is continuous and bounded by some $C<\infty$.*

Similar to the previous notation, we define $1_t(\gamma)\equiv 1\{q_t'\gamma>0\}$ while $1_t(\gamma,\bar{\gamma})\equiv 1\{q_t'\gamma\le 0<q_t'\bar{\gamma}\}$, which should not cause much confusion. Furthermore, we let the last element of q_t equal to -1.

Lemma I.1. *Let Assumption I.1 hold. Then, there exists $T_0<\infty$ such that for any $\vec{\gamma}$ in a neighbourhood of γ_0 , $K>0$ and for all $T>T_0$ and $\epsilon\ge T^{-1}$,*

$$\mathbb{P}\left\{\sup_{|\gamma-\vec{\gamma}|_2<\epsilon}\left|\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(z_t 1_t\left(\vec{\gamma},\gamma\right)-\mathbb{E}z_t 1_t\left(\vec{\gamma},\gamma\right)\right)\right|>K\right\}\le\frac{C}{K^4}\epsilon^2.$$

An obvious implication of this lemma is that when $\epsilon=a_T^{-1}$ for some sequence $a_T=O(T)$ the process in the display is $O_P\left(a_T^{-1/2}\right)$. It also leads to the following uniform bounds for empirical processes of mixing arrays.

Lemma I.2. *Let Assumption I.1 hold. For any $\eta>0$ and some $C,c>0$,*

$$\sup_{cT^{-1+2\varphi}\le|\gamma-\gamma_0|_2<C}\left[\left|\frac{1}{T^{1+\varphi}}\sum_{t=1}^{T}\left(z_t\left(1_t(\gamma)-1_t(\gamma_0)\right)-\mathbb{E}z_t\left(1_t(\gamma)-1_t(\gamma_0)\right)\right)\right|\right.$$
$$\left.-\eta T^{-2\varphi}\left|\gamma-\gamma_0\right|_2\right]\le O_P\left(\frac{1}{T}\right).$$

Lemma I.3. *Let Assumption I.1 hold. For any $\eta > 0$ and some $C, c > 0$,*

$$\sup_{cT^{-1+2\varphi} \le |\gamma-\gamma_0|_2 < C} \left[\left| \frac{1}{\sqrt{N}T^{1-\varphi}} \sum_{t=1}^{T} \left(z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) - \mathbb{E} z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) \right) \right| - \eta \left| \gamma - \gamma_0 \right|_2^2 \right]$$
$$\le O_P \left(\frac{1}{\left(NT^{1-2\varphi} \right)^{2/3}} \right).$$

We derive an extended continuous mapping theorem (CMT) in Lemma I.4, in the sense that we consider a transformation by a continuous stochastic process. This lemma extends Theorem 1.11.1 of van der Vaart and Wellner (1996) to allowing stochastic drifting functions $\mathbb{G}_n$ (while van der Vaart and Wellner (1996) requires $\mathbb{G}_n$ be deterministic).

Lemma I.4. *Suppose that as $n \to \infty$,*

$$\mathbb{G}_n(x) \Rightarrow \mathbb{G}(x)$$

over any compact set in $\mathbb{R}^m$, where $\mathbb{G}(\cdot)$ is a Gaussian process with continuous sample paths. Let f_n be a sequence of random functions from $\mathbb{R}^k$ onto $\mathbb{R}^m$ and assume that

$$f_n(z) \xrightarrow{P} f(z),$$

uniformly, where f is a deterministic function, and that for any $\eta > 0$ there exists $C_\eta < \infty$ such that

$$\mathbb{P}\left\{ \left| f_n(z) - f_n(z') \right|_2 > C_\eta \left| z - z' \right|_2 \text{ for all } z, z' \right\} < \eta,$$

for all n. Then,

$$\mathbb{G}_n(f_n(z)) \Rightarrow \mathbb{G}(f(z))$$

over any compact set.

I.1 Proofs of Lemmas

Proof of Lemma I.1. In this proof, c, C and so on denote generic constants. Let the dimension of q_t be denoted by $d_f = d + 1$ and partition $\gamma = (\psi', c)'$ and $q_t = (q_{1t}', -1)'$ Also let

$$J_T(\gamma) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(z_t 1_t(\vec{\gamma}, \gamma) - \mathbb{E} z_t 1_t(\vec{\gamma}, \gamma) \right).$$

First, note that Lemma 3.6 of Peligrad (1982) implies that there is a universal constant C, depending only on the ρ_m's, such that for any γ_1 and γ_2,

$$\begin{aligned} &\mathbb{E}\left|\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(z_t 1_t\left(\gamma_1,\gamma_2\right)-\mathbb{E}z_t 1_t\left(\gamma_1,\gamma_2\right)\right)\right|^4 \\ &\leq C\left(T^{-1}\mathbb{E}\left|z_t\right|^4\ 1_t\left(\gamma_1,\gamma_2\right)+\left(\mathbb{E}\left|z_t\right|^2\ 1_t\left(\gamma_1,\gamma_2\right)\right)^2\right). \end{aligned} \tag{I.1}$$

Consider $\gamma_1 = (\psi', c_1)'$ and $\gamma_2 = (\psi', c_2)'$, which are identical other than the last elements. Then,

$$1_t\left(\gamma_1,\gamma_2\right) = 1\left\{c_2 < q_{1t}'\psi \leq c_1\right\}$$

and thus there is a universal constant C such that

$$\begin{aligned} \mathbb{E}\left|z_t\right|^k\ 1_t\left(\gamma_1,\gamma_2\right) &= \mathbb{E}\left[\mathbb{E}\left(\left|z_t\right|^k\middle| q_t\right)1_t\left(\gamma_1,\gamma_2\right)\right] \\ &\leq C\mathbb{E}1_t\left(\gamma_1,\gamma_2\right) \leq C'\left|c_1-c_2\right| \end{aligned}$$

for $k = 2, 4$, as the densities of $q_t'\gamma$ are bounded uniformly. Thus, for any c_1, c_2 such that $|c_1 - c_2| \geq T^{-1}$,

$$\sup_{\psi}\mathbb{E}\left|\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(z_t 1_t\left(\gamma_1,\gamma_2\right)-\mathbb{E}z_t 1_t\left(\gamma_1,\gamma_2\right)\right)\right|^4 \leq C\left|c_1-c_2\right|^2. \tag{I.2}$$

Here, recall that ψ is the common element between γ_1 and γ_2.

Next, by Bickel and Wichura (1971), their equation (1), that

$$\sup_{\gamma}\left|J_T\left(\gamma\right)\right| \leq d\cdot M'' + \left|J_T\left(\widetilde{\gamma}\right)\right|,$$

where $\widetilde{\gamma}$ is the elementwise increament of $\vec{\gamma}$ by ϵ and the supremum is taken over a hyper cube $\{\gamma : 0 \leq \gamma_j - \vec{\gamma}_j \leq \epsilon, j = 1, ..., d\}$ and an upper bound for M'' is given by their Theorem 1. The precise definition of M'' is referred to Bickel and Wichura. It is sufficient to show that each of M'' and $|J_T(\widetilde{\gamma})|$ satisfies the conclusion of the lemma since $|a| + |b| > 2c$ implies that $|a| > c$ or $|b| > c$.

To apply their Theorem 1, we need to consider the increment of the process J_T around a block[4] $B = (\gamma_1, \gamma_2] = (\gamma_{12,}\gamma_{22}] \times \cdots \times (c_1, c_2]$ with each side of length greater than equal to

[4]It is sufficient to consider blocks with side length at least n^{-1} for the same reason as the remarks in the last paragraph in p. 1665.

T^{-1}, that is, consider

$$\begin{aligned}
J_T(B) &= \sum_{k_1=0,1}\cdots\sum_{k_{d+1}=0,1}(-1)^{d-k_1-\cdots-k_{d+1}} J_T(\gamma_{11}+k_1(\gamma_{21}-\gamma_{11}),...,c_1+k_{d+1}(c_2-c_1)) \\
&= \sum_{k_1=0,1}\cdots\sum_{k_d=0,1}(-1)^{d-k_1-\cdots-k_d} \\
&\quad\times(J_T(\gamma_{11}+k_1(\gamma_{21}-\gamma_{11}),...,c_1)-J_T(\gamma_{11}+k_1(\gamma_{21}-\gamma_{11}),...,c_2)).
\end{aligned}$$

Then, it follows from the c_r-inequality and (I.2) that for some $C, C', C'' < \infty$

$$\begin{aligned}
&\mathbb{E}|J_T(B)|^4 \\
\leq\ & C\sum_{k1=0,1}\cdots\sum_{k_d=0,1}\mathbb{E}|J_T(\gamma_{11}+k_1(\gamma_{21}-\gamma_{11}),...,c_1)-J_T(\gamma_{11}+k_1(\gamma_{21}-\gamma_{11}),...,c_2)|^4 \\
\leq\ & C'\sup_{\psi}\mathbb{E}\left|\frac{1}{T^{1/2}}\sum_{t=1}^{T}(z_t 1_t(\gamma_1,\gamma_2)-\mathbb{E}z_t 1_t(\gamma_1,\gamma_2))\right|^4, \text{ for } \gamma_j=(\psi', c_j), j=1,2 \\
\leq\ & C''|c_1-c_2|^2.
\end{aligned}$$

Now, without loss of generality we can assume that $\mu(B) \geq C'''|c_1-c_2|^d$, where μ denotes the Lebesque measure in $\mathbb{R}^d$, since we can derive the same bound by choosing the smallest side length of B as $c_2 - c_1$. This implies by the Cauchy-Schwarz inequality that their $\mathcal{C}(\beta,\gamma)$ condition holds with $\beta = 4$ and $\gamma = 2/d$, and thus, by their Theorem 1, we conclude

$$\mathbb{P}\left\{M'' > K\right\} \leq \frac{C}{K^4}\mu(T)^{2/d} \leq \frac{C}{K^4}\epsilon^2,$$

for some $C < \infty$.

Furthermore, the Markov inequality, the moment bound in (I.1), the boundedness of the density of $q_t'\gamma$ imply that

$$\mathbb{P}\{|J_T(\widetilde{\gamma})| > K\} \leq \frac{C}{K^4}\epsilon^2,$$

for some $C < \infty$. This completes the proof. ∎

Proof of Lemma I.2. Define $A_{T,j} = \{\theta : (j-1)T^{-1+2\varphi} \leq |\gamma-\gamma_0|_2 < jT^{-1+2\varphi}\}$ and

$$R_T^2 = T\sup_{T^{-1+2\varphi}<|\gamma-\gamma_0|\leq C}\left[|\mathbb{D}_T(\gamma)| - \eta T^{-2\varphi}|\gamma-\gamma_0|_2\right],$$

where $\mathbb{D}_T(\gamma) = \frac{1}{T^{1+\varphi}} \sum_{t=1}^{T} \left(z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) - \mathbb{E} z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) \right)$. Then, for any $m > 0$,

$$
\begin{aligned}
&\mathbb{P}\left\{R_T > m\right\} \\
&= \mathbb{P}\left\{T\left|\mathbb{D}_T(\gamma)\right| > \eta|\gamma-\gamma_0|T^{1-2\varphi} + m^2 \text{ for some } \gamma\right\} \\
&\leq \sum_{\ell=2}^{\infty} \mathbb{P}\left\{T\left|\mathbb{D}_T(\gamma)\right| > \eta(\ell-1) + m^2 \text{ for some } \gamma \in A_{T\ell}\right\} \\
&\leq C' \sum_{\ell=2}^{\infty} \frac{\ell^2}{\left(\eta(\ell-1)+m^2\right)^4},
\end{aligned}
$$

where the last equality is due to Lemma I.1 with $K = T^{-1/2+\varphi}\left(\eta(\ell-1)+m^2\right)$ and $\epsilon = \ell T^{-1+2\varphi}$. The last term is finite for any $\eta > 0$ and can be made arbitrarily small by choosing sufficiently large m, which completes the proof. ∎

Proof of Lemma I.3. Define $A_{T,j} = \{\gamma : (j-1) \leq \widetilde{n}^{2/3} \left|\gamma - \gamma_0\right|_2^2 < j\}$ with $\widetilde{n} = NT^{1-2\varphi}$ and

$$
R_T^2 = \widetilde{n}^{2/3} \sup_{T^{-1+2\varphi} < |\gamma-\gamma_0| \leq C} \left[\left|\mathbb{D}_T(\gamma)\right| - \eta\left|\gamma-\gamma_0\right|_2^2\right],
$$

where $\mathbb{D}_T(\gamma) = \frac{1}{\sqrt{N}T^{1-\varphi}} \sum_{t=1}^{T} \left(z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) - \mathbb{E} z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) \right)$. Then, for any $\varepsilon > 0$, we can find m such that

$$
\begin{aligned}
\mathbb{P}\left\{R_T > m\right\} &= \mathbb{P}\left\{\widetilde{n}^{2/3}\left|\mathbb{D}_T(\gamma)\right| > \eta\widetilde{n}^{2/3}|\gamma-\gamma_0|^2 + m^2 \text{ for some } \gamma\right\} \\
&\leq \sum_{\ell=2}^{\infty} \mathbb{P}\left\{\widetilde{n}^{2/3}\left|\mathbb{D}_T(\gamma)\right| > \eta(\ell-1) + m^2 \text{ for some } \gamma \in A_{T\ell}\right\} \\
&\leq C' \sum_{\ell=2}^{\infty} \frac{\widetilde{n}^{2/3}}{\left(\eta(\ell-1)+m^2\right)^4} \frac{\ell}{\widetilde{n}^{2/3}} \leq \varepsilon
\end{aligned}
$$

where the first and second inequalities follow from the union bound and Lemma I.1 with $K = \widetilde{n}^{-1/6}\left(\eta(\ell-1)+m^2\right)$ and $\epsilon = \sqrt{\frac{\ell}{\widetilde{n}^{2/3}}}$, respectively, and the third by choosing sufficiently large m. This completes the proof. ∎

Proof of Lemma I.4. First, we show the stochastic equicontinuity of $\mathbb{G}_n\left(f_n(z)\right)$. For any positive ε and η, there exist $\delta > 0$ and N such that for all $n > N$,

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{|z-z'|_2<\delta}\left|\mathbb{G}_n\left(f_n\left(z\right)\right)-\mathbb{G}_n\left(f_n\left(z'\right)\right)\right|_2>\eta\right\} \\
&\leq \mathbb{P}\Bigg\{\sup_{|z-z'|_2<\delta}\left|\mathbb{G}_n\left(f_n\left(z\right)\right)-\mathbb{G}_n\left(f_n\left(z'\right)\right)\right|_2>\eta \text{ and } \left|f_n\left(z\right)-f_n\left(z'\right)\right|_2\leq C\left|z-z'\right|_2 \\
&\qquad \text{and } \sup_z\left|f_n\left(z\right)\right|_2\leq C\Bigg\} \\
&\quad + \mathbb{P}\left\{\left|f_n\left(z\right)-f_n\left(z'\right)\right|_2>C\left|z-z'\right|_2\right\}+\mathbb{P}\left\{\sup_z\left|f_n\left(z\right)\right|_2>C\right\} \\
&\leq \mathbb{P}\left\{\sup_{|x-x'|_2<\delta/C}\left|\mathbb{G}_n\left(x\right)-\mathbb{G}_n\left(x'\right)\right|_2>\eta\right\}+\frac{\varepsilon}{2} \\
&\leq \varepsilon,
\end{aligned}$$

where the second inequality is due to the set inclusion and the given condition on f_n with boundedness of z and the last one follows from the stochastic equicontinuity of $\mathbb{G}_n$.

Second, for the fidi note that

$$\mathbb{G}_n\left(f_n\left(z\right)\right)-\mathbb{G}_n\left(f\left(z\right)\right)\xrightarrow{p}0$$

due to the stochastic equicontinuity of $\mathbb{G}_n$ as $f_n\left(z\right)\xrightarrow{p}f\left(z\right)$. Therefore, for any finite collection $\left(z_1,...,z_p\right)'$, $\left(\mathbb{G}_n\left(f_n\left(z_1\right)\right),...,\mathbb{G}_n\left(f_n\left(z_p\right)\right)\right)'=\left(\mathbb{G}_n\left(f\left(z_1\right)\right),...,\mathbb{G}_n\left(f\left(z_p\right)\right)\right)'+o_P\left(1\right)\xrightarrow{d}\left(\mathbb{G}\left(f\left(z_1\right)\right),...,\mathbb{G}\left(f\left(z_p\right)\right)\right)'$ due to the weak convergence of $\mathbb{G}_n$. ∎

J Additional Tables of Simulation Results

In this part of the appendix, we collect the additional tables of the simulation results that are omitted from the main text.

Table A-2: Simulation Results: Baseline Model ($T = N = 200$)

	Mean Bias	RMSE	Coverage
Scenario (i): Oracle			
β_1	-0.0025	0.0427	0.948
β_2	0.0015	0.0383	0.947
δ_1	0.0012	0.0749	0.962
δ_2	-0.0039	0.0678	0.959
Scenario (ii): Observed Factors/No Selection on g_t			
β_1	-0.0033	0.0430	0.943
β_2	0.0013	0.0385	0.942
δ_1	0.0042	0.0759	0.956
δ_2	-0.0027	0.0684	0.954
ϕ_2	0.0002	0.0655	
ϕ_4	-0.0011	0.0495	
Ave. Cor. Regime Prediction:			0.9929 (0.0074)
Scenario (iii): Observed Factors/Selection on g_t			
β_1	-0.0034	0.0431	0.943
β_2	0.0013	0.0385	0.940
δ_1	0.0045	0.0759	0.959
δ_2	-0.0027	0.0685	0.954
ϕ_2	-0.0053	0.0646	
ϕ_3	0.0010	0.0110	
ϕ_4	-0.0023	0.0526	
Ave. Cor. Regime Prediction:			0.9925 (0.0080)
Correct Factor Selection:			0.985
Scenario (iv): Unobserved Factors			
β_1	-0.0002	0.0435	0.945
β_2	0.0032	0.0391	0.940
δ_1	-0.0062	0.0795	0.952
δ_2	-0.0085	0.0702	0.957
γ_2	-0.0003	0.5098	
γ_3	-0.0061	0.4977	
γ_4	-0.0061	0.3784	
Ave. Cor. Regime Prediction:			0.9799 (0.0122)

Note: The average correct regime prediction (Ave. Cor. Regime Prediction) measures the average proportion such that the predicted regime of $1\{g_t'\widehat{\phi} > 0\}$ (or $1\{f_t'\widehat{\gamma} > 0\}$ in (iv)) is equal to the true regime of $1\{g_t'\phi_0 > 0\}$ (or $1\{f_t'\gamma_0 > 0\}$ in (iv)): $\widehat{E}\left(\frac{1}{T}\sum_{t=1}^{T} 1\left\{1\{g_t'\hat{\phi} > 0\} = 1\{g_t'\phi_0 > 0\}\right\}\right)$, where the expectation $\widehat{E}$ is taken over simulation draws.

Table A-3: Unobserved Factors with Different N Sizes

	Mean Bias	RMSE
$N = 100$		
β_1	0.0097	0.0473
β_2	0.0077	0.0407
δ_1	-0.0397	0.1015
δ_2	-0.0376	0.0939
γ_2/γ_1	0.0016	0.0802
Ave. Cor. Regime Prediction:	0.9741 (0.0133)	
$N = 200$		
β_1	0.0067	0.0462
β_2	0.0050	0.0386
δ_1	-0.0252	0.0966
δ_2	-0.0241	0.0850
γ_2/γ_1	-0.0014	0.0629
Ave. Cor. Regime Prediction:	0.9821 (0.0107)	
$N = 400$		
β_1	0.0038	0.0460
β_2	0.0028	0.0379
δ_1	-0.0129	0.0880
δ_2	-0.0142	0.0795
γ_2/γ_1	-0.0010	0.0500
Ave. Cor. Regime Prediction:	0.9870 (0.0087)	
$N = 1600$		
β_1	0.0010	0.0443
β_2	0.0006	0.0373
δ_1	-0.0029	0.0851
δ_2	-0.0056	0.0759
γ_2/γ_1	0.0011	0.0392
Ave. Cor. Regime Prediction:	0.9934 (0.0062)	

Note: See the note under Table A-2 for the definition of Ave. Cor. Regime Prediction.

Table A-4: Computation Time for Different Sample Sizes (unit=second)

	Algorithm	T=200	T=300	T=400	T=500
Min	MIQP	1.87	2.85	3.97	5.23
Median	MIQP	1.99	3.04	4.39	5.66
Mean	MIQP	1.99	3.09	4.34	5.66
Max	MIQP	2.21	3.69	4.73	6.07

Table A-5: Computation Time for Different Sizes of x_t (unit=second)

	Algorithm	$d_x = 1$	$d_x = 2$	$d_x = 3$	$d_x = 4$
Min	MIQP	1.87	2.16	2.39	2.46
Median	MIQP	1.99	2.31	2.52	2.76
Mean	MIQP	1.99	2.30	2.51	2.76
Max	MIQP	2.21	2.54	2.85	3.06

Table A-6: Computation Time for Different Sizes of g_t (unit=second)

	Algorithm	$d_g = 2$	$d_g = 3$	$d_g = 4$	$d_g = 5$
Min	MIQP	1.87	2.04	4.79	78.78
Median	MIQP	1.99	2.17	6.42	410.35
Mean	MIQP	1.99	2.18	6.56	445.15
Max	MIQP	2.21	2.38	9.68	1389.86

Table A-7: Computation Time of Large Models: $T = 500$

$T = 500$	Algorithm	$d_g = 6$	$d_g = 8$	$d_g = 10$
$d_x = 6$				
Min Time	BCD	611.34	610.69	610.42
	MIQP	1806.03	1806.08	1806.15
Median Time	BCD	612.35	611.61	611.76
	MIQP	1806.21	1806.18	1806.22
Max Time	BCD	614.57	612.63	612.34
	MIQP	1806.77	1807.28	1806.48
Median Obj.	BCD	0.23	0.24	0.24
	MIQP	0.25	0.3	0.29
Convergence	BCD	1	1	1
	MIQP	0	0	0
$d_x = 8$				
Min Time	BCD	611.70	611.75	611.97
	MIQP	1806.58	1806.59	1806.7
Median Time	BCD	613.49	612.03	640.21
	MIQP	1806.74	1806.74	1806.78
Max Time	BCD	614.59	612.69	1151.22
	MIQP	1807.32	1807.38	1807.51
Median Obj.	BCD	0.24	0.24	0.59
	MIQP	0.27	0.37	1.24
Convergence	BCD	1	1	1
	MIQP	0	0	0
$d_x = 10$				
Min Time	BCD	613.51	612.74	612.84
	MIQP	1807.77	1807.79	1807.77
Median Time	BCD	614.69	613.14	614.23
	MIQP	1807.84	1807.92	1807.84
Max Time	BCD	616.05	613.66	1693.54
	MIQP	1808.37	1808.32	1808.04
Median Obj.	BCD	0.24	0.23	0.27
	MIQP	0.27	0.69	1.42
Convergence	BCD	1	1	1
	MIQP	0	0	0

Table A-8: Computation Time of Large Models: $T = 1000$

$T = 1,000$	Algorithm	$d_g = 6$	$d_g = 8$	$d_g = 10$
$d_x = 6$				
Min Time	BCD	628.65	625.61	627.10
Median Time	BCD	635.58	645.17	634.25
Max Time	BCD	1344.56	1344.56	1163.16
Median Obj.	BCD	0.25	0.26	0.24
Convergence	BCD	1.00	1.00	1.00

References

ARCONES, M. A., AND B. YU (1994): "Central limit theorems for empirical andu-processes of stationary mixing sequences," *Journal of Theoretical Probability*, 7(1), 47–71.

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

BAI, J., AND S. NG (2008): "Forecasting economic time series using targeted predictors," *J. Econom*, 146(2), 304–317.

BERTSIMAS, D., A. KING, AND R. MAZUMDER (2016): "Best subset selection via a modern optimization lens," *Annals of Statistics*, 44(2), 813–852.

BICKEL, P. J., AND M. J. WICHURA (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications," *Annals of Mathematical Statistics*, 42(5), 1656–1670.

BILLINGSLEY, P. (1968): *Convergence of probability measures*, Wiley Series in probability and Mathematical Statistics: Tracts on probability and statistics. Wiley.

DAVIDSON, J. (1994): *Stochastic limit theory: An introduction for econometricians*. Oxford University Press, Oxford.

FAN, J., Y. LIAO, AND M. MINCHEVA (2013): "Large covariance estimation by thresholding principal orthogonal complements (with discussion)," *Journal of the Royal Statistical Society, Series B*, 75, 603–680.

GONÇALVES, S., AND B. PERRON (2019): "Bootstrapping factor models with cross sectional dependence," *J. Econom*, forthcoming.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. Academic Press, Boston,.

HANSEN, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64(2), 413–430.

LEE, S., M. H. SEO, AND Y. SHIN (2011): "Testing for threshold effects in regression models," *Journal of the American Statistical Association*, 106(493), 220–231.

PELIGRAD, M. (1982): "Invariance principle for mixing sequences of random variables," *Annals of Probability*, 10(4), 968–981.

POTTER, S. M. (1995): "A nonlinear approach to US GNP," *J. Appl. Econom.*, 10(2), 109–125.

VAN DER VAART, A., AND J. WELLNER (1996): *Weak convergence and empirical processes*. Springer, New York.