DEPARTMENT OF ECONOMICS
WORKING PAPER SERIES

2018-14

McMASTER UNIVERSITY

Department of Economics
Kenneth Taylor Hall 426
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

http://www.mcmaster.ca/economics/

Factor-Driven Two-Regime Regression*

Sokbae Lee† Columbia University

Yuan Liao‡ Rutgers University

Myung Hwan Seo§ Seoul National University

Youngki Shin¶ McMaster University

October 25, 2018

Abstract

We propose a novel two-regime regression model where the switching between the regimes is driven by a vector of possibly unobservable factors. When the factors are latent, we estimate them by the principal component analysis of a much larger panel data set. Our approach enriches conventional threshold models in that a vector of factors may represent economy-wide shocks more realistically than a scalar observed random variable. Estimating our model brings new challenges as well as opportunities in terms of both computation and asymptotic theory. We show that the optimization problem can be reformulated as mixed integer optimization and present two alternative computational algorithms. We derive the asymptotic distributions of the resulting estimators under the scheme that the threshold effect shrinks to zero. In particular, with latent factors, not only do we establish the conditions on factor estimation for a strong oracle property, which are different from those for smooth factor augmented models, but we also identify semi-strong and weak oracle cases and establish a phase transition that describes the effect of first stage factor estimation as the cross-sectional dimension of panel data increases relative to the time-series dimension. Moreover, we develop a consistent factor selection procedure with a penalty term on the number of factors and present a complementary bootstrap testing procedure for linearity with the aid of efficient computational algorithms. Finally, we illustrate our methods via Monte Carlo experiments and by applying them to factor-driven threshold autoregressive models of US macro data.

Keywords: threshold regression, factors, mixed integer optimization, panel data, phase transition, oracle properties, ℓ_0-penalization

*We would like to thank Don Andrews, Bruce Hansen, Zhongjun Qu and the seminar participants at Emory, Michigan State, NYU, Wisconsin-Madison, Northwestern, Yale, and 2018 ASSA Winter Meeting for helpful comments. We would like to thank the Social Sciences and Humanities Research Council of Canada (SSHRC-435-2018-0275) and the European Research Council for financial support (ERC-2014-CoG-646917-ROMIA) and the UK Economic and Social Research Council for research grant (ES/P008909/1) to the CeMMAP.

†Address: 420 West 118th Street, New York, NY 10027, USA. E-mail: sl3841@columbia.edu.

‡Address: 75 Hamilton St., New Brunswick, NJ 08901, USA. Email: yuan.liao@rutgers.edu.

§Address: 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. E-mail: myunghseo@snu.ac.kr.

¶Address: 1280 Main St. W., Hamiloton, ON L8S 4L8, Canada. Email: yshin12@gmail.com.

Contents

1 Introduction

In this paper, we consider a two-regime regression model. Suppose that the dependent variable y_t is generated from

$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{f_t'\gamma_0 > 0\} + \varepsilon_t, \tag{1.1}$$

$$\mathbb{E}\left(\varepsilon_t | \mathcal{F}_{t-1}\right) = 0,\ t = 1, \ldots, T, \tag{1.2}$$

where x_t and f_t are adapted to the filtration $\mathcal{F}_{t-1}$, $(\beta_0, \delta_0, \gamma_0)$ is a vector of unknown parameters, and the unobserved random variable ε_t satisfies the conditional mean restriction in (1.2). We interpret that f_t is a vector of "factors" determining regime switching. Note that when the single index $f_t'\gamma_0$ is strictly positive, the regression function becomes $x_t'(\beta_0 + \delta_0)$, and if $f_t'\gamma_0 \leq 0$, the regression function reduces to $x_t'\beta_0$. We allow for either observable or unobservable factors. For the latter, we assume that they can be recovered from a large panel of variables (see Section 6 for details). In light of this feature, we call the regression model in (1.1) and (1.2) a *factor-driven two-regime regression model.*

Our paper is closely related to the literature on threshold models with unknown change points.[1] In the conventional threshold regression model, an intercept term and a scalar observed random variable constitute f_t. For instance, Hansen (2000) studied the model in which $1\{f_t'\gamma_0 > 0\}$ in (1.1) is replaced by $1\{q_t > \widetilde{\gamma}_0\}$ for some observable scalar variable q_t with a scalar unknown parameter $\widetilde{\gamma}_0$. In real-world problems, it might be controversial to choose which observed variable plays the role of f_t. For example, if the two different regimes represent expansions and contractions in an economy, arguably it is difficult to single out one observed random variable that governs the business cycle. On the contrary, our proposed model introduces a regime change due to a single index of factors, thereby allowing us to model a regime switch based on a potentially large number of covariates.

To give the sense of our model, we consider Auerbach and Gorodnichenko (2012), Ramey and Zubairy (2018) and Tenreyro and Thwaites (2016) as empirical examples of threshold models in marcoeconomics.[2] In both Auerbach and Gorodnichenko (2012) and Ramey and Zubairy (2018), the key issue was the size of fiscal multipliers in recessions in the US. Tenreyro and Thwaites (2016) investigated whether the US economy responded differently to monetary policy shocks, depending on the state of the business cycle. Auerbach and Gorodnichenko (2012) and Tenreyro and Thwaites (2016) estimated smooth regime-switching models using a seven quarter moving average of the output growth rate as the threshold variable. Their primary results relied on a fixed level of intensity of regime switching. In Ramey and Zubairy (2018), the baseline results assume that the US economy is in a slack state if the unemploy-

[1] See, e.g., Caner and Hansen (2001), Chan (1993), Hansen (1996, 1999, 2000), Seijo and Sen (2011), Seo and Linton (2007) and Tong (1990) among many others.

[2] See also related work by Auerbach and Gorodnichenko (2013a,b) using data from OECD countries.

ment rate is above 6.5%. One important empirical question is how to choose a threshold index and how regimes are determined accordingly. Since economic activities take place in many heterogenous sectors with delayed feedback effects across sectors, it is debatable whether the modeling of regime switching in the aforementioned papers is realistic. Auerbach and Gorodnichenko (2012) remarked that the choice of the threshold variable "is not trivial because there is no clear-cut theoretical prescription for what this variable should be." To check the baseline results, Tenreyro and Thwaites (2016) examined robustness to changes in the intensity of regime switching and Ramey and Zubairy (2018) conducted various robustness checks using different thresholds.

We propose a systematic approach by estimating the change in the business cycle or economic slack using a vector of factors with unknown parameters. The latent index $f_t'\gamma_0$ includes the special case that a specific observed variable is above a threshold (e.g. by setting $f_t'\gamma_0 =$ unemployment rate at time t $- 6.5$) but is much more general. Especially, when the space of relevant covariates is large, it may be more appropriate to employ the framework of latent approximate factor models.[3] For instance, the diffusion indexes that are estimated from a large number of macroeconomic variables can be adopted as f_t (Stock and Watson, 2002a,b; Ludvigson and Ng, 2009). This allows the regimes to be potentially determined by a large number of economic variables and corresponds to the case of estimated factors. We focus on the case that the unobserved factors are estimated by the principal component analysis (PCA). In the empirical literature, factors typically enter into the regression model linearly (e.g. Bernanke, Boivin, and Eliasz, 2005). One notable exception is Galvão and Owyang (2018), who developed a Bayesian approach to estimate a factor-augmented smooth-transition vector autoregressive model. Unlike the literature, estimated factors enter into our regression model in (1.1) nonsmoothly, which effectuates a difficult estimation problem.

In addition to extending the threshold model to (1.1), we propose an ℓ_0-penalized consistent factor selection procedure to select the active factors that enter the index $f_t'\gamma_0$. For example, Ramey and Zubairy (2018) considered GDP gap and capacity utilization as alternative measures of economic slack and used them in their robustness check. Using our framework, we could consider unemployment, GDP gap, capacity utilization and their lags all together and select active factors in a data-dependent way, without settling on the choice of relevant factors on an ad hoc basis. Identifying the active factors would enable us to understand better which economic variables drive regime changes, thereby providing insights into the state-dependent effects of fiscal policy shocks.

In view of the conditional mean restriction in (1.2), a natural strategy to estimate $(\beta_0, \delta_0, \gamma_0)$ is to rely on least squares. A least squares estimator for our model brings out

[3]See, e.g., Bai (2003), Bai and Ng (2002, 2006), Fan, Liao, and Mincheva (2013), Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002a,b) among many others. See also Bai and Wang (2016) for a recent review on factor models in economics.

new challenges in terms of both computation and asymptotic theory. First of all, when the dimension of f_t is larger than 2, it is computationally demanding to estimate $(\beta_0, \delta_0, \gamma_0)$. We overcome this difficulty by developing new computational algorithms based on the method of mixed integer optimization (MIO). Specifically, we propose two alternative approaches that complement each other. Thanks to the developments in MIO solution algorithms and fast computing environments, the MIO has become increasingly used in recent applications. Well-known numerical solvers such as CPLEX and Gurobi can be used to effectively solve large-scale MIO problems. See, for example, Bertsimas, King, and Mazumder (2016, Section 2.1) for discussions on computational advances in solving the MIO problems.

Second, we establish asymptotic properties of our proposed estimator by adopting Hansen (2000)'s framework of a diminishing thresholding effect. That is, we assume that $\delta_0 = T^{-\varphi} d_0$ for some $\varphi > 0$ and a non-diminishing vector d_0. We focus on the region $\varphi \in (0, 1/2)$.[4] When the factors are latent, we assume that $T = O(N)$ throughout the paper, where N is the number of cross-sectional variables to construct the PCA estimates of f_t. It turns out that the asymptotic distribution for the estimator of $\alpha_0 \equiv (\beta_0', \delta_0')'$ is identical to that when γ_0 were known regardless of factors are directly observable or not; therefore, the estimator of α_0 enjoys an oracle property, provided that $T = O(N)$.

The issue is more complicated for the distribution of the estimator of γ_0. When factors are directly observable, we prove that

$$T^{1-2\varphi} \left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \operatorname*{argmin}_{g \in \mathcal{G}} B(g) + 2W(g),$$

where $B(g)$ represents a "drift function" of the criterion function, which is linear with a kink at zero, $W(g)$ is a mean-zero Gaussian process and $\mathcal{G}$ is a rescaled parameter space. However, when factors are not directly observable, the estimation error from the PCA plays an essential role and may slow down the rates of convergence, depending on the relation between N and T. Specifically, we show that

$$\left(\left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi}\right) \left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \operatorname*{argmin}_{g \in \mathcal{G}} A(\omega, g) + 2W(g),$$

with a new drift function $A(\omega, g)$ that depends on $\omega = \lim \sqrt{N} T^{-(1-2\varphi)} \in [0, \infty]$. On one hand, when $\omega = \infty$, we have that $A(\omega, g) = B(g)$, so the limiting distribution becomes the same as if the factors were observable. This case corresponds to the *super-consistency rate* as in Hansen (2000). On the other hand, when $\omega = 0$, it turns out that $A(\omega, g)$ is quadratic

[4]This region corresponds to strongly identified cases. The weakly identified case of $\varphi = 1/2$ is not dealt in this paper. Wang (2018) focused on the case of $\varphi = 1/2$ when the unknown threshold parameter is a scalar and there is only one observed random variable in f_t. His approach is not applicable in our model since it is based on re-ordering the observed scalar threshold variable. We also exclude the case of $\varphi = 0$, which requires a separate development of asymptotic theory.

in g, corresponding to a *cube root rate* similar to the maximum score estimator (Kim and Pollard, 1990). Furthermore, both the drift function and the resulting rates of convergence have continuous transitions as ω changes between 0 and ∞. Therefore, one of our key findings for the estimator of γ_0 is the occurrence of a *phase transition* from a *weak-oracle* limiting distribution to a *semi-strong* oracle one and then to a *strong* oracle one as ω increases.

The remainder of the paper is organized as follows. Section 2 provides sufficient conditions under which $(\beta_0', \delta_0', \gamma_0')'$ is identified. In Section 3, we propose the least squares estimator and two complementary algorithms to compute the proposed estimator. In Section 4, we establish asymptotic theory when f_t is directly observed. In Section 5, we propose a variable selection procedure for active factors and prove its consistency. In Section 6, we consider estimation when f_t is a vector of latent factors, propose two-step estimators via the method of principal components, and analyze asymptotic properties of our proposed estimators. In Section 7, we consider inference and focus on testing the linearity of the regression model in (1.1). Section 8 gives the results of Monte Carlo experiments. In Section 9, we illustrate our methods by applying them to threshold autoregressive models of US GNP and unemployment.[5] Section 10 concludes, and the appendices provides details that are omitted from the main text.

1.1 Notation

The sample size is denoted by T and the transpose of a matrix is denoted by a prime. The true parameter is denoted by the subscript 0, whereas a generic element is without the subscript. For example, γ is an element of the parameter space Γ and γ_0 is the true parameter. The Euclidean norm is denoted by $|\cdot|_2$, the Frobenius norm of a matrix is denoted by $|\cdot|_F$, and the ℓ_0-norm is denoted by $|\cdot|_0$. For a generic random variable or vector z_t, let its density function be denoted by p_{z_t}. Similarly, let $p_{y_t|x_t}(y)$ denote the conditional density of y_t given x_t for the random vectors y_t and x_t. Abbreviation *a.s.* refers to almost surely.

2 Identification

We first establish 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.

We make the convention that the constant 1 is the first element of x_t and -1 is the last element of f_t. Define $Z_t(\gamma) = (x_t', x_t' 1\{f_t'\gamma > 0\})'$ and $\alpha = (\beta', \delta')'$. Then, we can rewrite the

[5]The replication R codes for both the Monte Carlo experiments and empirical applications are available at `https://github.com/yshin12/fadtwo`.

model as

$$y_t = Z_t(\gamma_0)' \alpha_0 + \varepsilon_t.$$

Note that since only the sign of the index $f_t'\gamma_0$ determines the regime switching, the scale of γ_0 is not identifiable. We consider two kinds of scale normalization for γ_0. Formally, we make the following assumption on the parameter space for (α_0, γ_0). Let d_x and d_f denote the dimensions of x_t and f_t, respectively.

Assumption 1 (Parameter Space)**.** $\alpha_0 \in \mathbb{R}^{2d_x}$ *and* $\gamma_0 \in \Gamma \equiv \{(1, \gamma_2')' : \gamma_2 \in \Gamma_2\}$*, where* $\Gamma_2 \subset \mathbb{R}^{d_f - 1}$ *is a compact set.*

If there is no random variable in f_t with a non-zero coefficient, γ_0 is unidentifiable. Assumption 1 avoids this directly by assuming that the first coefficient of γ_0 is 1.[6] 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 2.1 (Alternative Scale Normalization)**.** We may consider an alternative parameter space for γ_0: $\gamma_0 \in \Gamma \equiv \{\gamma : |\gamma|_2 = 1, \gamma \neq (0, ..., 0, 1)'$, 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 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.

In view of the conditional mean zero restriction in (1.2), it is natural to establish conditions under which both α_0 and γ_0 are identified by the L_2-loss. Introduce the excess loss

$$R(\alpha, \gamma) \equiv \mathbb{E}(y_t - x_t'\beta - x_t'\delta 1\{f_t'\gamma > 0\})^2 - \mathbb{E}(y_t - x_t'\beta_0 - x_t'\delta_0 1\{f_t'\gamma_0 > 0\})^2. \tag{2.1}$$

Note that $R(\alpha_0, \gamma_0) = 0$. In order to establish that $R(\alpha_0, \gamma_0) > 0$ whenever $(\alpha, \gamma) \neq (\alpha_0, \gamma_0)$, we make the following regularity conditions.

Assumption 2 (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 $\gamma \neq \gamma_0$,

$$\mathbb{E}\left[\left(x_t'\delta_0\right)^2 1\{B_{\gamma t}\}\right] > 0. \tag{2.2}$$

[6] 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.

(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,*

$$\inf_{\gamma\in\Gamma} \mathbb{E}\left[x_t x_t' 1\{A_{1\gamma t}\}\right] > 0 \quad \textit{and} \quad \inf_{\gamma\in\Gamma} \mathbb{E}\left[x_t x_t' 1\{A_{2\gamma t}\}\right] > 0. \tag{2.3}$$

Note that under Assumption 2(i), $R(\cdot,\cdot)$ is continuous. The condition (2.2) ensures the presence of a change in the regression function. If $\delta_0 = 0$, then (2.2) is not satisfied. A sufficient condition for (2.2) 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 (2.3) is satisfied, for example, if

$$\mathbb{E}\left[x_t x_t' 1\left\{\inf_{\gamma\in\Gamma} f_t'\gamma > 0\right\}\right] > 0 \text{ and } \mathbb{E}\left[x_t x_t' 1\left\{\sup_{\gamma\in\Gamma} f_t'\gamma \leq 0\right\}\right] > 0. \tag{2.4}$$

Note that (2.4) requires that (i) the parameter space Γ satisfies

$$\mathbb{P}\left(\bigcap_{\gamma\in\Gamma}\{f_t'\gamma > 0\}\right) > 0 \text{ and } \mathbb{P}\left(\bigcap_{\gamma\in\Gamma}\{f_t'\gamma \leq 0\}\right) > 0 \tag{2.5}$$

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 2.2. It is possible to provide sufficient conditions for Assumption 2 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 2.1 (Identification)**.** *If Assumptions 1 and 2 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 2.1 gives the basis for our estimator given in the next section.

3 Least Squares Estimator via Mixed Integer Optimization

We now propose the least squares estimator and two complementary algorithms to compute the proposed estimator. For the computational purpose, we assume that $\alpha\in\mathcal{A}\subset\mathbb{R}^{2d_x}$ for some known compact set $\mathcal{A}$. In practice, one can take a large $2d_x$-dimensional hyperrectangle so that the resulting estimator is not on the boundary of $\mathcal{A}$.

Theorem 2.1 suggests that the unknown parameters can be estimated by the least squares[7]:

$$(\widehat{\alpha},\widehat{\gamma})=\arg\min_{(\alpha',\gamma')'\in\mathcal{A}\times\Gamma}\mathbb{S}_T(\alpha,\gamma) \tag{3.1}$$

$$\text{subject to:}\quad \tau_1\le\frac{1}{T}\sum_{t=1}^{T}1\{f_t'\gamma>0\}\le\tau_2, \tag{3.2}$$

where

$$\mathbb{S}_T(\alpha,\gamma)\equiv\frac{1}{T}\sum_{t=1}^{T}(y_t-x_t'\beta-x_t'\delta 1\{f_t'\gamma>0\})^2. \tag{3.3}$$

We assume that the restriction (3.2) is satisfied when $\gamma=\gamma_0$ almost surely. Here $0<\tau_1<\tau_2<1$ for some predetermined τ_1 and τ_2 (e.g. $\tau_1=0.05$ and $\tau_2=0.95$). In the special case that $1\{f_t'\gamma_0>0\}=1\{q_t>\widetilde{\gamma}_0\}$ with a scalar variable q_t and a parameter $\widetilde{\gamma}_0$, it is standard to assume that the parameter space for $\widetilde{\gamma}_0$ is between the τ and $(1-\tau)$ quantiles of q_t for some known $0<\tau<1$. We can interpret (3.2) as a natural generalization of this type of restriction so that the proportion of one regime is never too close to 0 or 1.

When γ is of high dimension, the naive grid search would not work well. We overcome this computational difficulty by replacing the naive grid search with mixed integer optimization (MIO). We present two alternative classes of MIO algorithms below.

3.1 A Joint Approach

Our first algorithm is based on mixed integer quadratic programming (MIQP), which jointly estimates (α,γ) and is guaranteed to obtain a global solution once it is found. To write the

[7]The estimate of γ_0 is not unique because the objective function in (3.3) involves indicator functions. However, we can work with any estimator that solves (3.1) under (3.2) since the resulting sample splitting of two regimes and the estimates of α_0 would be uniquely determined. The same issue already exists for the case of a scalar threshold variable.

original least squares problem in MIQP, we introduce $d_t = 1\{f_t'\gamma > 0\}$, and $\ell_{j,t} = \delta_j d_t$ for $j = 1, \ldots, d_x, t = 1, \ldots, T$, where δ_j denote the j-th element of δ. Then the least squares objective function can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - \sum_{j=1}^{d_x} x_{j,t}\ell_{j,t}\right)^2, \tag{3.4}$$

which is a quadratic function and minimized with respect to $(\beta, \delta, d_1, \ldots, d_T, \ell_{1,1}, ..., \ell_{d_x,T})$, subject to $\ell_{j,t} = \delta_j d_t$ for all (j,t) and additional constraints to be presented below. Observe that (3.4) adds new integer variables $d_1, \ldots, d_T$, each taking value in $\{0, 1\}$.

The goal is to introduce only linear constraints with respect to $(\beta, \delta, d_1, \ldots, d_T, \ell_{1,1}, ..., \ell_{d_x,T})$, and reach an MIQP that is equivalent to the original least squares problem, so that we can apply modern MIO packages (e.g. Gurobi) to solve MIQP. First note that the assumption $\alpha \in \mathcal{A}$ implies that there exist known upper and lower bounds for δ_j: $L_j \le \delta_j \le U_j$. In addition, to make sure that $\ell_{j,t} = \delta_j d_t$ for each j and t, impose two additional restrictions: $d_t L_j \le \ell_{j,t} \le d_t U_j$ and $L_j(1 - d_t) \le \delta_j - \ell_{j,t} \le U_j(1 - d_t)$. It is then straightforward to check these constraints imply $\ell_{j,t} = \delta_j d_t$. To introduce another key constraint, define

$$M_t \equiv \max_{\gamma \in \Gamma} |f_t'\gamma|$$

for each $t = 1, \ldots, T$, where Γ is the parameter space for γ_0. One can compute M_t easily for each t using linear programming. We store them as inputs to our algorithm. The following new constraints ensure that the reformulated problem (3.4) is the same as the original problem:

$$(d_t - 1)(M_t + \epsilon) < f_t'\gamma \le d_t M_t,$$

where $\epsilon > 0$ is a small predetermined constant (e.g. $\epsilon = 10^{-6}$).

The following defines an algorithm for joint estimation.

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

$$\min_{\beta,\delta,\gamma,\boldsymbol{d},\boldsymbol{\ell}} \mathbb{Q}_T(\beta, \boldsymbol{\ell}) \equiv \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - \sum_{j=1}^{d_x} x_{j,t}\ell_{j,t}\right)^2 \tag{3.5}$$

subject to

$$
\begin{aligned}
&(\beta, \delta) \in \mathcal{A},\ \gamma \in \Gamma, \\
&L_j \leq \delta_j \leq U_j, \\
&(d_t - 1)(M_t + \epsilon) < f_t'\gamma \leq d_t M_t, \\
&d_t \in \{0, 1\}, \\
&d_t L_j \leq \ell_{j,t} \leq d_t U_j, \\
&L_j(1 - d_t) \leq \delta_j - \ell_{j,t} \leq U_j(1 - d_t), \\
&\tau_1 \leq \frac{1}{T}\sum_{t=1}^{T} d_t \leq \tau_2
\end{aligned}
\tag{3.6}
$$

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

Our proposed algorithm is mathematically equivalent to the original least squares problem (3.1) subject to (3.2) in terms of values of objective functions. Formally, we state it as the following theorem.

Theorem 3.1. *Let* $(\bar{\alpha}, \bar{\gamma})$ *denote a solution to the joint optimization problem using MIQP described above. For all* $\epsilon > 0$, $\mathbb{S}_T(\widehat{\alpha}, \widehat{\gamma}) = \mathbb{S}_T(\bar{\alpha}, \bar{\gamma})$, *where* $(\widehat{\alpha}, \widehat{\gamma})$ *is a solution to* (3.1) *subject to* (3.2).

The proposed algorithm in Section 3.1 may run slowly when the dimension d_x of x_t is large. To mitigate this problem, we reformulate the joint optimization in Appendix B.2 and use the alternative formulation in our numerical work; however, we present a simpler form here to help readers follow our basic ideas more easily.

3.2 An Iterative Approach

While the MIQP jointly estimates (α, γ) and aims at obtaining a global solution, it may not compute as fast as necessary in large scale problems. To mitigate the issue of scalibility, we introduce a faster alternative approach based on mixed integer linear programming (MILP), whose objective function is linear in d_t. The algorithm solves for α and γ iteratively, starting with an initial value that can be obtained through a crude grid search. At step k, given $\widehat{\alpha}^{k-1}$ that is obtained in the previous step, we estimate γ by solving

$$
\min_{\gamma \in \Gamma, d_1, \ldots, d_T} \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\widehat{\beta}^{k-1} - x_t'\widehat{\delta}^{k-1} d_t\right)^2 \tag{3.7}
$$

subject to similar constraints as in the joint approach. The following defines an algorithm for the iterative estimation.

[Iterative Estimation]

1. (Grid Construction) Construct a grid, say $\Gamma_T \equiv \{\gamma_j\}_{j=1}^{m_T}$, of Γ, such that $\max_{\gamma\in\Gamma}\min_j |\gamma - \gamma_j| \to 0$ as $T \to \infty$.

2. (Initial Joint Estimation) For the given grid Γ_T, obtain the initial estimate

$$\left(\widehat{\alpha}^0, \widehat{\gamma}^0\right) = \underset{\alpha\in\mathcal{A},\gamma\in\Gamma_T}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T}\left(y_t - Z_t(\gamma)'\alpha\right)^2.$$

3. Iterate the following steps (a)-(c), beginning with $k = 1$ and terminating at a prespecified number $\bar{K}$.

   (a) For the given $\widehat{\alpha}^{k-1}$, obtain an estimate $\widehat{\gamma}^k$ via the mixed integer linear optimization algorithm

$$\min_{\gamma\in\Gamma,d_1,\ldots,d_T} \frac{1}{T}\sum_{t=1}^{T}\left\{(x_t'\widehat{\delta}^{k-1})^2 - 2(y_t - x_t'\widehat{\beta}^{k-1})x_t'\widehat{\delta}^{k-1}\right\} d_t \tag{3.8}$$

   subject to

$$\begin{aligned} &(d_t - 1)(M_t + \epsilon) < f_t'\gamma \le d_t M_t, \\ &d_t \in \{0,1\} \quad \text{for each } t = 1,\ldots,T, \\ &\tau_1 \le \frac{1}{T}\sum_{t=1}^{T} d_t \le \tau_2. \end{aligned} \tag{3.9}$$

   (b) For the given $\widehat{\gamma}^k$, obtain

$$\widehat{\alpha}^k = \left[\frac{1}{T}\sum_{t=1}^{T} Z_t\left(\widehat{\gamma}^k\right) Z_t\left(\widehat{\gamma}^k\right)'\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} Z_t\left(\widehat{\gamma}^k\right) y_t$$

   (c) Let $k = k + 1$.

Note that the least squares problem (3.7) is equivalent to (3.8) due to the fact that $d_t^2 = d_t$. Therefore, the objective function in (3.8) is linear in d_t.

The iterative approach is generally faster than the joint approach since first, it is easier to solve an MILP problem than to solve an MIQP problem and second, $\widehat{\alpha}^k$ has an explicit solution. We also note that the specification of Γ_T in step 1 for the initial grid search can be crude. Our theoretical study shows that the algorithm works well as long as the initial value is consistent for γ_0. We provide weak conditions on the grid Γ_T and $\bar{K}$ under which the algorithm produces asymptotically equivalent solutions to the joint approach after only a few iterations. More specifically, when factors are known, $\bar{K}=1$ is sufficient; when factors are unknown and estimated, $\bar{K}=2$ iterations would suffice.

4 Asymptotic Properties with Known Factors

We split asymptotic properties of the estimator into two cases of known and unknown factors. In this section, we consider the former.

Assumption 3. *(i) $\{x_t, f_t, \varepsilon_t\}$ is a sequence of strictly stationary, ergodic, and ρ-mixing with $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$, $\mathbb{E}\left|x_t\right|_2^4<\infty$, and there exists a constant $C<\infty$ such that $\mathbb{E}\left(\left|x_t\right|_2^8\left|f_t'\gamma=0\right.\right)<C$ and $\mathbb{E}\left(\varepsilon_t^8|f_t'\gamma=0\right)<C$ for all $\gamma\in\Gamma$.*

(ii) $\{\varepsilon_t\}$ is a martingale difference sequence.

(iii) The smallest eigenvalue of $\mathbb{E}\left[Z_t\left(\gamma\right)Z_t\left(\gamma\right)'\right]$ is bounded away from zero for all $\gamma\in\Gamma$.

Assumption 4 (Diminishing jump). *(i) For some $0<\varphi<1/2$ and $d_0\neq 0$, assume $\delta_0=d_0T^{-\varphi}$.*

(ii) The conditional density $p_{u_t|f_{2t}}(u)$ of $u_t:=f_t'\gamma_0$ given f_{2t}, $\mathbb{E}\left[\left(x_t'd_0\right)^2|f_{2t},u_t=u\right]$ and $\mathbb{E}\left[\left(\varepsilon_t x_t'd_0\right)^2|f_{2t},u_t=u\right]$ are continuous and bounded away from zero at $u=0$ a.s.

(iii) For some $M<\infty$, $\inf_{|r|_2=1}\mathbb{E}\left(|f_{2t}'r|\,1\left\{|f_{2t}|_2\leq M\right\}\right)>0$.

Most of conditions in Assumptions 3 and 4 are a natural extension of the standard conditions in the literature. See e.g. Hansen (2000) when $f_t=(q_t,-1)'$ for a scalar random variable. A few conditions merit comments. In particular, Assumption 4 (iii) is a rank condition on f_{2t} due to the vector of threshold parameter to be estimated and it is in terms of the first moment because of the asymptotic linear approximation of criterion function near γ_0. It also allows for discrete variables in f_{2t}. Observe that the condition that $\mathbb{E}\left(\left|x_t\right|_2^8\left|f_t'\gamma=0\right.\right)<C$ for all $\gamma\in\Gamma$ does not necessarily imply that $\mathbb{E}\left|x_t\right|_2^4<\infty$ since the conditional expectation in the former is restricted to the event $f_t'\gamma=0$. Assumption 4 (ii) ensures the presence of a jump, not just a kink at the change point.

Theorem 4.1. *Let $\mathcal{G}:=\{g\in\mathbb{R}^{d_f}:g_1=0\}$. Let Assumptions 1, 2, 3, and 4 hold. Assume further that α_0 is in the interior of $\mathcal{A}$ and γ_0 is in the interior of Γ. In addition, let W denote a mean-zero Gaussian process whose covariance kernel is given by*

$$H\left(s,g\right):=\frac{1}{2}\mathbb{E}\left[\left(\varepsilon_t x_t'd_0\right)^2\left(\left|f_t'g\right|+\left|f_t's\right|-\left|f_t'\left(g-s\right)\right|\right)p_{u_t|f_{2t}}(0)\right]. \tag{4.1}$$

Then (i) as $T \to \infty$, for the estimators $\widehat{\alpha}$ and $\widehat{\gamma}$ obtained via the joint approach, we have that

$$\sqrt{T}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, (\mathbb{E}Z_t(\gamma_0)Z_t(\gamma_0)')^{-1}\text{var}(Z_t(\gamma_0)\varepsilon_t)(\mathbb{E}Z_t(\gamma_0)Z_t(\gamma_0)')^{-1}),$$
$$T^{1-2\varphi}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \underset{g \in \mathcal{G}}{\text{argmin}}\, \mathbb{E}\left[(x_t'd_0)^2 \left|f_t'g\right| p_{u_t|f_{2t}}(0)\right] + 2W(g),$$

and $\sqrt{T}(\widehat{\alpha} - \alpha_0)$ and $T^{1-2\varphi}(\widehat{\gamma} - \gamma_0)$ are asymptotically independent.
(ii) The iterative estimators, $\widehat{\alpha}^k$ and $\widehat{\gamma}^k$, have the identical asymptotic distribution as $(\widehat{\alpha}, \widehat{\gamma})$, for any finite $k \geq 1$, provided that the grid $\Gamma_T \equiv \{\gamma_j\}_{j=1}^{m_T}$ of Γ satisfies $\max_{\gamma \in \Gamma} \min_j |\gamma - \gamma_j| \to 0$ as $T \to \infty$.

The proof of Theorem 4.1 is given in Appendix C, along with proofs of consistency and rates of convergence. Note that the normalization scheme is embedded in the asymptotic distribution. Since $\gamma_1 = 1$, the minimum in the limit is taken after fixing the first element of g at zero (recall that $\mathcal{G} = \{g \in \mathbb{R}^{d_f} : g_1 = 0\}$).

Remark 4.1. The limiting Gaussian process $W(g)$ becomes the two-sided Brownian motion when $f_t = (q_t, -1)'$ and $\gamma_0 = (1, \widetilde{\gamma}_0)'$. To see this, note that $|f_t'g| + |f_t's| - |f_t'(g-s)| = |\widetilde{g}| + |\widetilde{s}| - |\widetilde{g} - \widetilde{s}|$ since in this case, $f_t'g = -\widetilde{g}$ for $g_1 = 0$. In addition, note that $|\widetilde{g}| + |\widetilde{s}| - |\widetilde{g} - \widetilde{s}| = 0$ if $\widetilde{g}$ and $\widetilde{s}$ are of opposite signs. The resulting limiting distribution of $\widehat{\gamma}$ then becomes the one derived by Hansen (2000).

Remark 4.2. Seo and Linton (2007) consider a similar model with known factors; however, they propose a smoothed estimator of (α_0, γ_0) and focus on the case of $\varphi = 0$. They show that their estimator of γ_0 is asymptotically normal at a rate slower than $1/T$. The asymptotic distribution result for α_0 in Theorem 4.1 is the same as that of Seo and Linton (2007), whereas the asymptotic result for γ_0 is different.

5 Selecting Relevant Factors

We consider factor selection with known factors. In applications, it is often difficult to have *a priori* knowledge regarding which variables constitute f_t in (1.1). Suppose that there are potentially a 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; however, this uncertainty can be easily adopted in our framework with the help of MIO.

To be specific, decompose $f_t = (f_{1t}', f_{2t}', -1)'$,[8] 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 $p = |f_{2t}|_0$, where

[8]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.

$|\cdot|_0$ denotes the ℓ_0-norm. 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 $j = 1, \ldots, p$. 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. We carry out factor selection and joint estimation by adopting an ℓ_0-penalized approach.

For a given penalty parameter $\lambda > 0$, define

$$\begin{aligned} \widetilde{\gamma} = & \arg\min_{\gamma\in\Gamma} \min_{\beta,\delta} \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - x_t'\delta 1\left\{f_t'\gamma > 0\right\}\right)^2 + \lambda|\gamma|_0 \\ & \text{subject to (3.2).} \end{aligned} \tag{5.1}$$

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

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

$$\min_{\beta,\delta,\gamma,\boldsymbol{d},\boldsymbol{\ell},\boldsymbol{e}} \widetilde{\mathbb{Q}}_T(\beta,\boldsymbol{\ell}) \equiv \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - \sum_{j=1} x_{j,t}\ell_{j,t}\right)^2 + \lambda\sum_{m=1}^{p} e_m \tag{5.2}$$

subject to (3.6) and

$$\begin{aligned} & e_m\underline{\gamma_2} \le \gamma_{2m} \le e_m\overline{\gamma_2}, \\ & \underline{p} \le \sum_{m=1}^{p} e_m \le \overline{p}, \\ & e_m \in \{0,1\} \quad \text{for each } m = 1, \ldots, p. \end{aligned} \tag{5.3}$$

Finally, re-estimate the model using only selected factors via the method given in Section 3.1.

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 ($\sum_{m=1}^{p} e_m$). The choice of λ is crucial to establish factor selection consistency. We provide formal theory below.

Theorem 5.1. *Let $S(\gamma) = \{j : \gamma_j \neq 0\}$ and $S_0 = S(\gamma_0)$. Let Assumptions 1, 2, 3, and 4 hold. Suppose that $\lambda \to 0$, $\lambda T \to \infty$, and p is fixed. Then,*

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

Theorem 5.1 states that our factor selection procedure is consistent, provided that $\lambda T \to \infty$. In Appendix D.1, we provide a detailed description of the relevant factor selection procedure when unknown parameters are estimated based on an iterative approach and establish conditions under which the iterative procedure yields consistent factor selection. In the paper, we consider active factor selection only for observed factors. When factors are unobservable but estimated via the PCA, interpretation of each estimated factor is more involved since factors are identified only up to some random rotation. Furthermore, a small number of estimated factors are typically used in applications—dimension reduction is already achieved via factor estimation—hence, there is relatively less demand for active factor selection in this case.

6 Estimation with Unobserved Factors

In this section, we consider the case that the factors are estimated. In empirical macroeconomics and finance, it is common to have many cross-sectional random variables (N) that are associated with the time series data of interest (y_t). In this paper, we focus on the PCA estimator of the unobserved factors.

6.1 The Model

Consider the following factor model:

$$\mathcal{Y}_t = \Lambda g_{1t} + e_t,\ t = 1, \ldots, T, \tag{6.1}$$

where $\mathcal{Y}_t$ is an $N \times 1$ vector of time series, Λ is an $N \times K$ matrix of factor loadings, g_{1t} is a $K \times 1$ vector of common factors, and e_t is an $N \times 1$ vector of idiosyncratic components. Throughout this section, we make it explicit that there is a constant term in the factors and replace the regression model in (1.1) with

$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{g_t'\phi_0 > 0\} + \varepsilon_t, \tag{6.2}$$

where $g_t = (g_{1t}', -1)'$ is a vector of unknown factors in (6.1) plus a constant term (-1) and ϕ_0 is a vector of unknown parameters.[9] In addition, we allow g_{1t} to contain lagged (dynamic) factors, but we treat them as static factors and estimate them using the PCA without losing the validity of the estimated factors. Likewise, g_t can embed the threshold structure as in our equation for y_t.

It is well known that g_t is identifiable and estimable by the PCA up to an invertible matrix transformation, say $H_T' g_t$, whose exact form will be given in Section 6.6. Therefore, it

[9] Recall that d_f is the dimension of f_t. Hence, $d_f = K + 1$ in this section.

is customary in the literature (see, e.g., Bai (2003) and Bai and Ng (2006)) to treat $H_T' g_t$ as a centering object in the limiting distribution of estimated factors. Following this convention, in this section, let

$$f_t := H_T' g_t \quad \text{and} \quad \gamma_0 := H_T^{-1}\phi_0. \tag{6.3}$$

Using the fact that $g_t'\phi_0 = f_t'\gamma_0$, we can rewrite (6.2) as the original formulation in (1.1):

$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{f_t'\gamma_0 > 0\} + \varepsilon_t.$$

Hence, in this section, γ_0 depends on the sample, but we suppress dependence on T for the sake of notational simplicity.

6.2 Two-Step Estimation

Our estimation procedure now consists of two steps: in the first step, a $(K+1)\times 1$ vector of estimated factors and the constant term, say $\widetilde{f}_t$, are obtained by the method of principal components. In the second step, unknown parameters (α_0, γ_0) are estimated with $\widetilde{f}_t$ as inputs.

To describe estimated factors, let $\mathcal{Y}$ be the $T \times N$ matrix whose t-th row is $\mathcal{Y}_t'$. Let $(\widetilde{f}_{11}, \ldots, \widetilde{f}_{1T})$ be the $K \times T$ matrix, whose rows are K eigenvectors (multiplied by $\sqrt{T}$) associated with the largest K eigenvalues of $\mathcal{Y}\mathcal{Y}'/NT$ in decreasing order. In the second step, the unknown parameters are estimated by

$$(\widehat{\alpha}, \widehat{\gamma}) = \underset{(\alpha',\gamma')' \in \mathbb{R}^{2d_x}\times\Gamma}{\operatorname{argmin}} \widetilde{\mathbb{S}}_T(\alpha, \gamma)$$

$$\text{subject to:} \quad \tau_1 \le \frac{1}{T}\sum_{t=1}^{T} 1\{\widetilde{f}_t'\gamma > 0\} \le \tau_2,$$

where

$$\widetilde{\mathbb{S}}_T(\alpha, \gamma) \equiv \frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\beta - x_t'\delta 1\{\widetilde{f}_t'\gamma > 0\})^2 \tag{6.4}$$

and $\widetilde{f}_t \equiv (\widetilde{f}_{1t}', -1)'$. Recall that we fix the normalization by Assumption 1; that is, the first element of γ is fixed at 1.[10] The algorithm for computing $(\widehat{\alpha}, \widehat{\gamma})$ is the same as in Section 3.

[10] One caveat of this normalization scheme is that the sign of the first element of f_t may not be the same as that of the first element of g_t due to random rotation H_T; however, if we assume that $\delta_0 \neq 0$ and we also know the sign of one of non-zero coefficients of δ_0, we can determine the sign of the first element of f_t after estimating the model. This is a "labelling" problem that is common in models with hidden regimes. For simplicity, we assume that the first element of γ_0 is 1.

6.3 Regularity Conditions

We introduce assumptions needed for asymptotic results with estimated factors. We first replace Assumptions 1-4 with the following assumption. Define

$$\Phi_T := \{\phi : \phi = H_T \gamma \text{ for some } \gamma \in \Gamma_\epsilon\}, \tag{6.5}$$

where Γ_ϵ is an ϵ-enlargement of Γ.[11] The space Φ_T for ϕ is defined through H_T and excludes the case that $g_t'\phi$ is degenerate. The ϵ-enlargement of Γ is needed since the factors are latent.

Assumption 5. *(i) Assumptions 1, 2 and 4 (i) hold after replacing f_t and γ_0 with g_t and ϕ_0, respectively.*

(ii) $\{x_t, g_t, e_t, \varepsilon_t\}$ is a sequence of strictly stationary, ergodic, and ρ-mixing with $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$, and there exists a constant $C < \infty$ such that $\mathbb{E}(|x_t|_2^8 | g_t, e_t) < C$, $\mathbb{E}(\varepsilon_t^8 | g_t, e_t) < C$ a.s., and $g_t'\phi$ has a density that is continuous and bounded by C for all $\phi \in \Phi_T$.

Recall that in Assumption 3(i), we have assumed that there exists a constant C such that $\mathbb{E}\left(|x_t|_2^8 \,| f_t'\gamma = 0\right) < C$ and $\mathbb{E}\left(\varepsilon_t^8 | f_t'\gamma = 0\right) < C$ for all $\gamma \in \Gamma$. We strengthen this assumption to Assumption 5(ii) that requires that the 8th moments of $|x_t|_2$ and ε_t be almost surely bounded conditional on g_t and e_t.

The following assumption is needed to deal with estimated factors.

Assumption 6. *(i) $\lim_{N\to\infty} \frac{1}{N}\Lambda'\Lambda = \Sigma_\Lambda$ for some $K \times K$ matrix Σ_Λ, whose eigenvalues are bounded away from both zero and infinity.*

(ii) The eigenvalues of $\Sigma_\Lambda^{1/2}\mathbb{E}(g_{1t}g_{1t}')\Sigma_\Lambda^{1/2}$ are distinct.

(iii) All the eigenvalues of the $N \times N$ covariance $\mathrm{var}(e_t)$ are bounded away from both zero and infinity.

(iv) For any t, $\frac{1}{N}\sum_{s=1}^{T}\sum_{i=1}^{N} |\mathbb{E}e_{it}e_{is}| < C_\sigma$ for some $C_\sigma > 0$.

All four conditions are standard in the literature. Condition (iv) allows weak serial correlation among e_t.

To state further conditions, define λ_i' to be the i-th row of Λ, so that $\Lambda = (\lambda_1, ..., \lambda_N)'$. Define

$$\xi_{s,t} \;:=\; \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{is}e_{it} - \mathbb{E}e_{is}e_{it}),$$

[11] Note that ϕ cannot be a vector whose first K elements are zeros due to the normalization on γ and the block diagonal structure of H_T that will be defined in (6.7).

$$\eta_t := \frac{1}{\sqrt{TN}} \sum_{s=1}^{T} \sum_{i=1}^{N} g_{1s}(e_{is}e_{it} - \mathbb{E}e_{is}e_{it}),$$
$$\psi := \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} g_t e_{it} \lambda_i',$$
$$\zeta_t := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{it} e_{it}.$$

We require the following additional exponential-tail conditions.

Assumption 7. *There exist finite, positive constants C, C_1 and c_1 such that for any $x > 0$,*

$$\mathbb{P}(|\varpi|_2 > x) \leq C \exp(-C_1 x^{c_1}),$$

where ϖ is any of the following: $e_{it}, g_{1t}, \xi_{s,t}, \zeta_t, vec(\psi)$, and η_t.

These conditions impose exponential tail conditions on various terms. To explain this assumption, first note that it requires weak cross-sectional correlations among e_{it}. Furthermore, these terms are standardized sums of weakly dependent sequences, whose tails should be well behaved due to the Bernstein type inequality as in e.g. Merlevède, Peligrad, and Rio (2011). While these quantities are often assumed to have finite moments in the literature, these moment bounds would no longer be sufficient in the current context. Instead, exponential-type probability bounds are more useful for us to characterize the effect of the estimated factors. To understand this, note that under the regularity conditions in this section, we have the following asymptotic expansion:

$$\widetilde{f}_t = \widehat{f}_t + r_t, \quad \widehat{f}_t := H_T'(g_t + \frac{1}{\sqrt{N}} h_t), \tag{6.6}$$

where r_t is a remainder term,

$$H_T' := \begin{pmatrix} \widetilde{H}_T' & 0 \\ 0 & 1 \end{pmatrix}, \quad h_t := \begin{pmatrix} h_{1t} \\ 0 \end{pmatrix}, \quad h_{1t} := (\frac{1}{N}\Lambda'\Lambda)^{-1} \frac{1}{\sqrt{N}} \Lambda' e_t,$$

and the exact form of $\widetilde{H}_T$ is given below in (6.7). The diagonality in H_T and the zero element in h_t reflect the inclusion of the constant in $\widehat{f}_t$.[12] Because the estimated factors appear in the model highly nonlinearly and nonsmoothly through the indicator functions, it is necessary to establish the following uniform approximation result: uniformly for γ over a compact set,

$$\max_{t \leq T} \left| \mathbb{P}(\widetilde{f}_t'\gamma > 0) - \mathbb{P}(\widehat{f}_t'\gamma > 0) \right| \leq O\left(\frac{(\log T)^c}{T} \right) + \max_{t \leq T} \mathbb{P}\left(|r_t| > C \frac{(\log T)^c}{T} \right)$$

[12] By the same token, we could include other directly observable factors at the expense of more complicated notation. For the sake of clarity, we stick to the current setup.

for some constants $C, c > 0$. The above exponential-tail assumption then enables us to derive a sharp bound so that $\max_{t\leq T}\mathbb{P}(|r_t| > C(\log T)^c T^{-1})$ is asymptotically negligible.

In addition, to analyze the effect of the estimated factors in the leading term $\widehat{f}_t$, it is necessary to make assumptions on g_t and h_t that appear in (6.6). We do so in the following assumptions.

Assumption 8. *Let $u_t := g_t'\phi_0$. Let $\mathcal{Z}_t$ be a sequence of Gaussian random variables whose conditional distribution given (x_t, g_t) is $\mathcal{N}(0, \sigma^2_{h,x_t,g_t})$ with*

$$\sigma^2_{h,x_t,g_t} := \operatorname*{plim}_{N\to\infty} \mathbb{E}[(h_t'\phi_0)^2|x_t, g_t].$$

Furthermore, $\mathcal{Z}_t$ satisfies the following:

(i) As $N \to \infty$, $\sup_{x_t,g_t} |\mathbb{P}(h_t'\phi_0 < 0|x_t, g_t) - \mathbb{P}(\mathcal{Z}_t < 0|x_t, g_t)| = O(N^{-1/2})$.

(ii) There are positive constants c and C such that

$$\sup_{x_t,g_t}\sup_{|z|<c} p_{h_t'\phi_0|g_t,x_t}(z) < C,$$

$$\sup_{x_t,g_t}\sup_{|z|<c} |p_{h_t'\phi_0|g_t,x_t}(z) - p_{\mathcal{Z}_t|g_t,x_t}(z)| = o(1).$$

(iii) For some $c_0 > 0$, $\sigma^2_{h,x_t,g_t} > c_0$ a.s.

Assumption 8 requires that the conditional probability of $h_t'\phi_0 < 0$ and the conditional density of $h_t'\phi_0$ be uniformly approximated by the normal probability and the normal density in a neighborhood of zero. Since h_t is a cross-sectional average multiplied by $\sqrt{N}$, it converges in distribution to a normal random vector by the conventional cross-sectional central limit theorem (CLT). The rate $N^{-1/2}$ in condition (i) is a reminiscent of the Berry-Essen theorem, which is not stringent. Recall that the criterion function is not smooth and thus some conditional moment characterizes the asymptotic distribution. Assumption 8 strengthens the usual unconditional CLT to a conditional version in this spirit. Specifically, for some function $\Psi(\cdot)$ such that $\mathbb{E}|\Psi(x_t, g_t)| < \infty$, it ensures:

$$\mathbb{E}\left[\Psi(x_t, g_t)\left(1\{h_t'\phi_0 \leq 0\} - 1\{\mathcal{Z}_t \leq 0\}\right)\middle| x_t, g_t\right] = O(N^{-1/2}).$$

In the next assumption, recall that by the identification condition, we can write $\gamma = (1, \gamma_2)$, where 1 is the first element of γ. Correspondingly, let f_{2t} and $\widehat{f}_{2t}$ be the subvectors of f_t and $\widehat{f}_t$, excluding their first elements. Also, recall that $u_t = g_t'\phi_0 = f_t'\gamma_0$ and let $\breve{g}_t := g_t' + h_t/\sqrt{N}$.

Assumption 9. *There exist positive constants c, c_0, M_0 and M such that the following holds almost surely:*

(i) $\inf_{|u|<c} p_{\widehat{f}_t'\gamma_0|\widehat{f}_{2t},x_t}(u) \geq c_0$ *and* $\sup_{|f|_2<M_0} p_{f_{2t}|h_t}(f) < M$.

(ii) $\inf_{|u|<c} p_{u_t|f_{2t},h_t,x_t}(u) \geq c_0$. *For all* $|u_1| < c, |u_2| < c$,

$$|p_{u_t|h_t'\phi_0,f_{2t},x_t}(u_1) - p_{u_t|h_t'\phi_0,f_{2t},x_t}(u_2)| \leq M|u_1 - u_2|.$$

(iii) $\inf_{|r|_2=1} \mathbb{E}\left[|f_{2t}'r|^k 1\{|f_{2t}|_2 < M_0\}\right] \geq c_0$ *for* $k = 1, 2$.

(iv) $\sup_{|r|_2=1} \sup_{|u|<c} p_{g_t'r|h_t}(u) \leq M$.

(v) Each of $\inf_{\phi\in\Phi_T} |g_t'\phi|$, $\inf_{\phi\in\Phi_T} |\breve{g}_t'\phi|$, $\sup_{\phi\in\Phi_T} |h_t'\phi|$, *and* $\breve{g}_t'\phi_0$ *has a density function bounded and continuous at zero, where* Φ_T *is defined in* (6.5).

(vi) $\mathbb{E}\left[(x_t'd_0)^2 |g_t, h_t\right]$ *is bounded above by* M_0 *and below by* c_0.

(vii) For any s *and* w *that are linearly independent of* ϕ_0, $\mathbb{E}\left((\varepsilon_t x_t'd_0)^2|\breve{g}_t'\phi_0 = u, \breve{g}_t's, \breve{g}_t'w\right)$ *and* $p_{\breve{g}_t'\phi_0|\breve{g}_t's,\breve{g}_t'w}(u)$ *are continuously differentiable at* $u = 0$ *with bounded derivatives. Furthermore,* $\mathbb{E}\left((\varepsilon_t x_t'd_0)^4 |\breve{g}_t|_2^2 |\breve{g}_t'\phi_0\right) \leq M$.

Assumption 9 (i) and (iii) with $k = 1$ are required for the case $T^{2-4\varphi} = o(N)$, and conditions (ii)-(iv) are for the case $N = O(T^{2-4\varphi})$. Conditions (v) and (vi) are needed to expand the loss function. Condition (vii) is required for deriving the asymptotic distribution. These conditions control the local characteristics of the centered least squares criterion function near the true parameter value. Since the model is perturbed by the error in the estimated factors $\widehat{f}_t$ up to negligible approximation error, the centered criterion is a drifitng sequence. Its leading term changes depending on whether $N = O(T^{2-4\varphi})$ or N is bigger than that. The lower bounds in the above assumption are part of rank conditions that ensure that the leading terms are well-defined. As a result, it entails a phase transition on the distribution of $\widehat{\gamma}$. Since they are rather technical, we provide a more detailed discussion on Assumption 9 in Appendix E.2.

6.4 Rates of Convergence

The following theorem presents the rates of convergence for the estimators.

Theorem 6.1. *Let Assumptions 5-9 hold. Suppose* $T = O(N)$. *Then*

$$|\widehat{\alpha} - \alpha_0|_2 = O_P\left(\frac{1}{\sqrt{T}}\right)$$

and

$$|\widehat{\gamma}-\gamma_0|_2 = O_P\left(\frac{1}{T^{1-2\varphi}}+\frac{1}{(NT^{1-2\varphi})^{1/3}}\right).$$

Theorem 6.1 establishes conditions under which $\widehat{\alpha}$ converges to α_0 at the rate of $1/\sqrt{T}$ and $\widehat{\gamma}$ converges to γ_0 at different rates, depending on how N diverges to infinity relative to T. The convergence rate of $\widehat{\gamma}$ merits further explanation. First of all, when N is relatively large so that $T^{2-4\varphi}=o(N)$, $\widehat{\gamma}$ converges in probability to γ_0 at a super-consistent rate of $T^{-(1-2\varphi)}$. Contrary to this case, when N is relatively small in the sense that $N=o(T^{2-4\varphi})$, the estimated threshold parameter has a cube root rate:

$$|\widehat{\gamma}-\gamma_0|_2 \leq O_P\left(\frac{1}{(NT^{1-2\varphi})^{1/3}}\right),$$

which is similar to that of the maximum score type estimators (Kim and Pollard (1990)). Therefore, as $\sqrt{N}/T^{1-2\varphi}$ varies in $[0,\infty]$, the rate of convergence varies between the super-consistency rate of the usual threshold models to the cube root rate of the maximum score type estimators. Furthermore, the convergence rates exhibit a continuous transition from one to the other.

To explain this continuous transition phenomenon, we can show that uniformly in (α,γ), the objective function has the following expansion: there are functions $R_1(\cdot)$ and $R_2(\cdot,\cdot)$ such that

$$\widetilde{\mathbb{S}}_T(\alpha,\gamma)-\widetilde{\mathbb{S}}_T(\alpha_0,\gamma_0)=R_1(\gamma)+R_2(\alpha,\gamma).$$

To deal with non-smooth objective functions, a key step of the analysis is to derive a sharp lower bound for $R_1(\gamma)$. We now describe the form of such a lower bound. When N is relatively large, the effect of estimating latent factors is negligible, and $R_1(\gamma)$ has a high degree of non-smoothness. Similar to the usual threshold model, we have

$$R_1(\gamma)\geq CT^{-2\varphi}|\gamma-\gamma_0|_2-O_P(T^{-1}).$$

This lower bound leads to a super-consistency rate. On the other hand, when N is relatively small, there are extra noises arising from the cross-sectional idiosyncratic errors when estimating the latent factors, which we call "cross-sectional noises". A remarkable feature of our model is that the cross-sectional noises help "smooth" the objective function in this case. As a result, the behavior of $R_1(\gamma)$ is similar to that of the maximum score type estimators, where a quadratic lower bound can be derived:

$$R_1(\gamma)\geq CT^{-2\varphi}\sqrt{N}|\gamma-\gamma_0|_2^2-O_P(T^{-2\varphi}N^{-5/6}).$$

The quadratic lower bound together with a larger error rate then leads to a cube root rate type of convergence. See Lemmas E.5 and E.6 in the appendix for more details.

6.5 Consistency of Regime-Classification

We introduce an error rate in (in-sample) regime-classification:

$$\widehat{R}_T = \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widetilde{f}_t'\widehat{\gamma} > 0\right\} - 1\left\{g_t'\phi_0 > 0\right\}\right|.$$

Here, $1\{g_t'\phi_0 > 0\}$ is the true regime indicator, which is estimated by $1\left\{\widetilde{f}_t'\widehat{\gamma} > 0\right\}$; thus $\left|1\left\{\widetilde{f}_t'\widehat{\gamma} > 0\right\} - 1\left\{g_t'\phi_0 > 0\right\}\right|$ equals zero when the regime is correctly estimated. The uncertainty about the regime classification comes from either the estimation of the factors g_t or the parameter estimation $\widehat{\gamma}$ or both. We establish its convergence rate in the following theorem.

Theorem 6.2. *Let Assumptions 5-9 hold. Suppose $T = O(N)$. Then*

$$\widehat{R}_T = O_P\left(\left(NT^{1-2\varphi}\right)^{-1/3} + T^{-1+2\varphi} + N^{-1/2}\right).$$

This is a useful corollary of the derivation of the rates of convergence for the threshold estimator. If we observe the factors directly or if N is sufficiently large to produce the oracle estimate of g_t, the error in the regime-classification is of the same magnitude as that of the threshold estimate, $O_P\left(T^{-1+2\varphi}\right)$. Thus, we expect an excellent performance of our regime classification rule even with a moderate size of T.

6.6 Asymptotic Distribution

To describe the asymptotic distribution, we introduce additional notation. Let V_T denote the $K \times K$ diagonal matrix whose elements are the K largest eigenvalues of $\mathcal{Y}\mathcal{Y}'/NT$. Define

$$\widetilde{H}_T' := V_T^{-1}\frac{1}{T}\sum_{t=1}^{T}\widetilde{f}_{1t}g_{1t}'\frac{1}{N}\Lambda'\Lambda, \quad H_T := \operatorname{diag}(\widetilde{H}_T, 1), \quad \text{and} \quad H := \operatorname*{plim}_{T,N\to\infty} H_T, \tag{6.7}$$

where H is well defined, following Bai (2003). Let

$$\omega := \lim_{N,T\to\infty}\frac{\sqrt{N}}{T^{1-2\varphi}} \in [0,\infty], \quad \zeta_\omega := \max\{\omega, \omega^{1/3}\}, \quad \text{and} \quad M_\omega := \max\{1, \omega^{-1/3}\}.$$

Define

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

for $\omega \in (0, \infty]$, with the convention that $1/\omega = 0$ for $\omega = \infty$, and

$$A(0,g) \quad = \quad \mathbb{E}\left[(x_t'd_0)^2(g_t'Hg)^2 \middle| u_t = 0, \mathcal{Z}_t = 0 \right] p_{u_t,\mathcal{Z}_t}(0,0)$$

for $\omega = 0$. Define $\mathbb{Z}_t(\phi) := (x_t', x_t'1\{g_t'\phi > 0\})'$.

Theorem 6.3. *Let Assumptions 5-9 hold. Suppose $T = O(N)$. Let $\mathcal{G} := \{0\} \times \mathbb{R}^K$. In addition, let W denote a mean-zero Gaussian process whose covariance kernel is given by*

$$H_W(s,g) := \frac{p_{u_t}(0)}{2}\mathbb{E}\left[\left(\varepsilon_t x_t'd_0\right)^2 \left(|g_t'Hg| + |g_t'Hs| - |g_t'H(g-s)| \right) \middle| u_t = 0 \right].$$

Then, (i) for the joint estimators, as $N, T \to \infty$,

$$\sqrt{T}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}\left(0, \left(\mathbb{E}\mathbb{Z}_t(\phi_0)\mathbb{Z}_t(\phi_0)'\right)^{-1} \mathbb{E}\left(\mathbb{Z}_t(\phi_0)\mathbb{Z}_t(\phi_0)'\varepsilon_t^2\right) \left(\mathbb{E}\mathbb{Z}_t(\phi_0)\mathbb{Z}_t(\phi_0)'\right)^{-1}\right),$$
$$\left(\left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi}\right)(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \operatorname*{argmin}_{g\in\mathcal{G}} A(\omega, g) + 2W(g),$$

and $\sqrt{T}(\widehat{\alpha}-\alpha_0)$ and $\left(\left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi}\right)(\widehat{\gamma} - \gamma_0)$ are asymptotically independent. Moreover,

$$A(0,g) = \lim_{w\to 0} A(w,g).$$

(ii) The iterative estimators, $\widehat{\alpha}^k$ and $\widehat{\gamma}^k$, have the identical asymptotic distribution as $(\widehat{\alpha}, \widehat{\gamma})$, for any finite $k \geq 2$, provided that the grid $\Gamma_T \equiv \{\gamma_j\}_{j=1}^{m_T}$ of Γ satisfies $\max_{\gamma\in\Gamma}\min_j |\gamma - \gamma_j| \to 0$ as $T \to \infty$.

In the literature, Bai and Ng (2006, 2008) have shown that the oracle property (with regard to the estimation of the factors) holds for the linear regression if $T^{1/2} = o(N)$ and for the extremum estimation and GMM estimation if $T^{5/8} = o(N)$, when the estimated factors are included in the model. Thus, it appears that the oracle property demands a larger N as the nonlinearity of the estimating equation rises. In view of this, we regard our condition, $T = O(N)$, not too stringent since we need to deal with estimated factors inside the indicator functions.

Theorem 6.3 has shown that the relative size of N over T affects the shape of the limiting criterion function. We categorize the results into three groups. In all three cases, the reults enjoy certain oracle property.

- Strong Oracle: $T^{2-4\varphi} = o(N)$ or $\omega = \infty$. The drift function $A(\infty, g)$ is approximated by a linear function with a kink at $g = 0$. Intuitively, a bigger N makes the estimated factors $\widehat{f}_t$ more precise. This yields the oracle result for both $\widehat{\gamma}$ and $\widehat{\alpha}$. In particular, it is straightforward to show that $\left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi} = T^{1-2\varphi}$ and $A\left(\infty, g\right) = \mathbb{E}\left[\left(x_t' d_0\right)^2 \left|f_t' g\right| p_{u_t|f_{2t}}(0)\right] = \mathbb{E}\left[\left(x_t' d_0\right)^2 \left|f_t' g\right| \left|u_t = 0\right.\right] p_{u_t}(0)$, yielding the same rate and asymptotic distribution as in the known factor case.

- Weak Oracle: $N = o\left(T^{2-4\varphi}\right)$ or $\omega = 0$. The drift function $A(0, g)$ is approximated by a quadratic function in g (adjusted by $\sqrt{N}T^{-2\varphi}$) in a neighborhood of γ_0. Certainly, it is harder to identify the minimum when the function is quadratic, making itself smooth at the minimum, than when it has a kink at the minimum. This results in the change of the asymptotic distribution as well as the slower rate of convergence for $\widehat{\gamma}$ to $\left(NT^{1-2\varphi}\right)^{-1/3}$. However, the oracle property for $\widehat{\alpha}$ is preserved.

- Semi-Strong Oracle: $N \asymp T^{2-4\varphi}$ or $\omega \in (0, \infty)$. In this case, $A(\omega, g)$ has a continuous transition between the two extreme cases discussed above. The effect of estimating factors is non-negligible for $\widehat{\gamma}$ and yet the estimator enjoys the same rate of convergence. The estimator $\widehat{\alpha}$ continues to be oracle efficient.

Remark 6.1. It is worthwhile to note that $A\left(\omega, g\right)$ is continuous everywhere and $A\left(\omega, g\right) \to +\infty$ as $|g| \to +\infty$ for any ω. The continuity of $A\left(\omega, g\right)$ in ω for any g implies that the distribution of the argmin of the limit processes $A\left(\omega, g\right) + 2W\left(g\right)$ is also continuous in ω in virtue of the argmax continuous mapping theorem (see e.g. van der Vaart and Wellner, 1996).

Remark 6.2. The asymptotic distribution of $\widehat{\gamma}$ is well-defined for any ω. Specifically, the argmin of the limit Gaussian process is $O_P\left(1\right)$ since $A\left(\omega, g\right)$ is a deterministic function of order at least $|g|$ for any ω while the variance of $W\left(g\right)$ grows at the rate of $|g|$ as $g \to \infty$. Furthermore, it possesses a unique minimizer almost surely due to Lemma 2.6 of Kim and Pollard (1990) since the variance of $W\left(g\right) - W\left(s\right)$ is nonzero for any $g \neq s$ as shown in the proof.

Remark 6.3. In the case of observable factors, as shown in Theorem 4.1, $k \geq 1$ suffices for the iterative estimators, while in the case of estimated factors, as shown in the above theorem, $k \geq 2$. A careful examination of our proofs reveals that in the estimated factor case, $k = 1$ iteration only leads to a preliminary rate of convergence for $\widehat{\gamma} - \gamma_0 = O_P(T^{-1(1-2\varphi)} + N^{-1/2})$, which is sharp and leads to a proper limiting distribution only when $T^{2-4\varphi} = o(N)$. In the more general rate of N, however, we need one more iteration to ensure sharp asymptotic results.

6.7 Discussion of $A(\omega, g)$ and its Graphical Representation

We now present an alternative and more explicit exposition of $A(\omega, g)$. Let $p(\cdot)$ denote the density function of the standard normal and recall that $u_t = f_t'\gamma_0$. Then, for $\omega \in [0,1]$,

$$A(\omega, g) = 2\mathbb{E}\left[\left(x_t'd_0\right)^2 \int_0^{|f_t'g|} \left(|f_t'g| - x\right) p\left(\frac{\omega^{1/3}x}{\sigma_{h,x_t,g_t}}\right) dx \middle| u_t = 0\right],$$

and, for $\omega \in [1, \infty]$,

$$A(\omega, g) = 2\mathbb{E}\left[\left(x_t'd_0\right)^2 \int_0^{\omega|f_t'g|} \left(|f_t'g| - \frac{x}{\omega}\right) p\left(\frac{x}{\sigma_{h,x_t,g_t}}\right) dx \middle| u_t = 0\right]$$

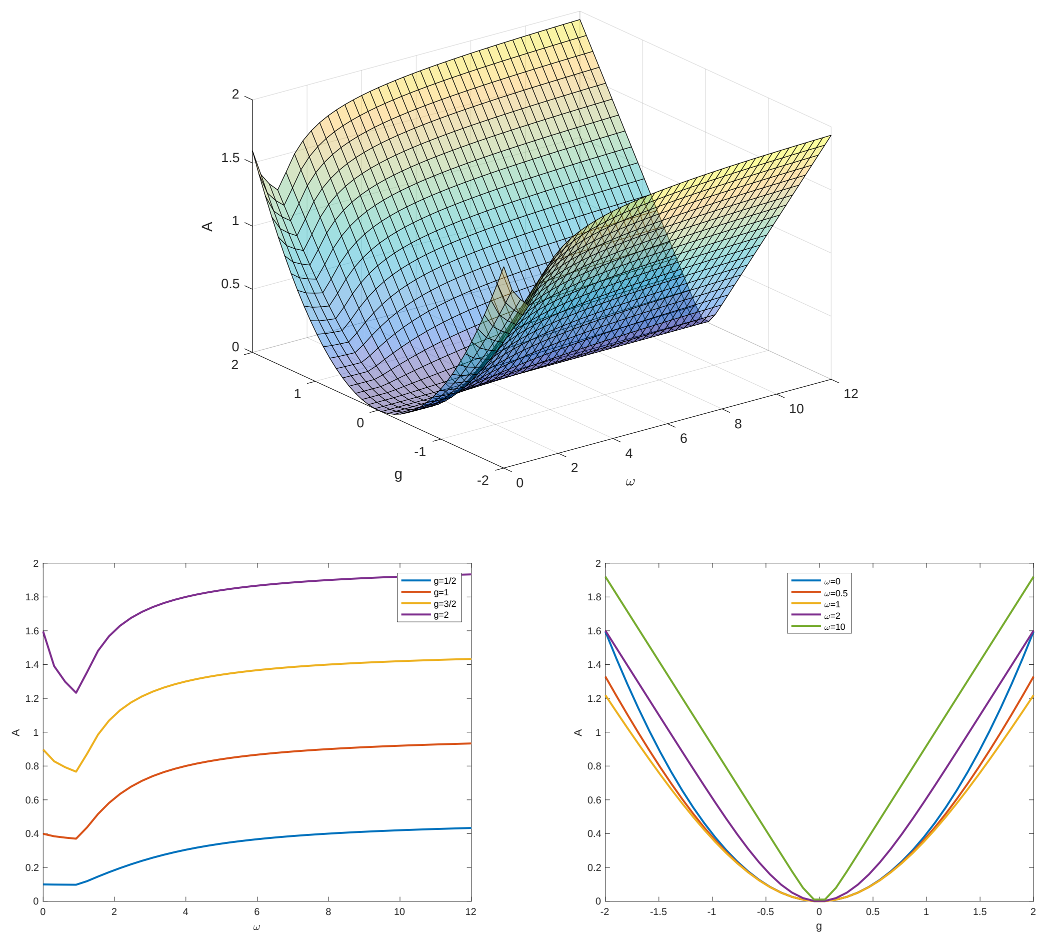
with the convention that $x/\omega = 0$ for $\omega = \infty$. This highlights the functional forms for $\omega = 0$ and $\omega = \infty$ and the presence of a possible kink at $\omega = 1$. Recall that $g_t'Hg = f_{2t}'g_2$ due to the normalization that $\gamma_{01} = 1$. Therefore, the conditional expectation in $A(\omega, g)$ does not degenerate.

To plot $A(\omega, g)$, we consider the simple case that $g_t = (q_t, -1)'$, $g = (0, g_2)'$, $x_t = 1$, $d_0 = 1$, and h_t and q_t are independent of each other. We simply write $g_2 = g$ for simplicity. The top panel of Figure 1 shows the three-dimensional graph of $A(\omega, g)$ and the bottom panel depicts the profile of $A(\omega, g)$ as a function of ω for several values of g and that of $A(\omega, g)$ as a function of g for given values of ω. First of all, it can be seen that $A(\omega, g)$ is continuous everywhere but has a kink at $\omega = 1$. As ω approaches zero, the shape of $A(\omega, g)$ is clearly quadratic in g; whereas as ω gets larger, it becomes almost linear in g. Also, note that $A(\omega, g)$ is quite flat around its minimum at $g = 0$ when ω is close to zero; however, $A(\omega, g)$ has a sharp minimum at zero for a larger value of ω. This reflects the fact that the rate of convergence increases as ω gets larger.

6.8 Phase Transition

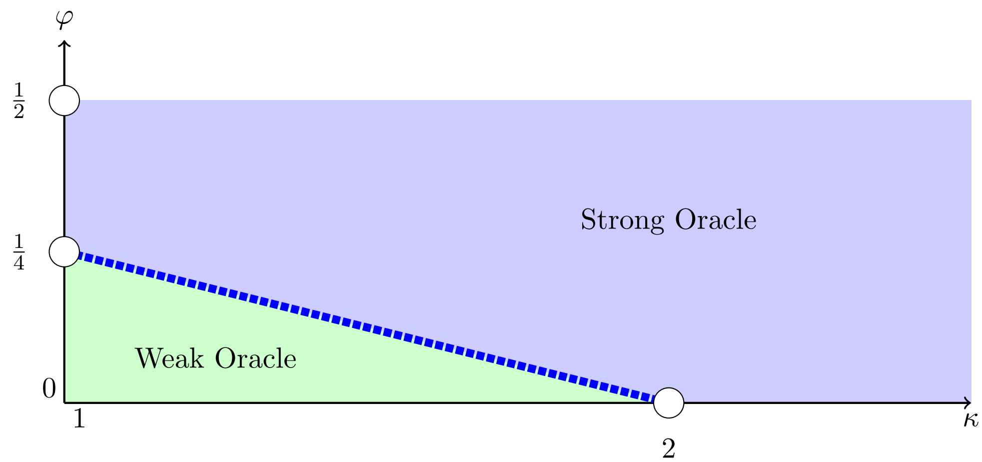
To demonstrate that our asymptotic results are sharp, we consider a special case that $N = T^\kappa$ for $\kappa \geq 1$. In this case, the asymptotic results can be depicted on the (κ, φ)-space. When $T^{1-2\varphi}$ diverges to infinity at a rate slower than $\left(NT^{1-2\varphi}\right)^{1/3} = T^{(\kappa+1-2\varphi)/3}$, the resulting convergence rates and asymptotic distributions for $\widehat{\gamma}$ and $\widehat{\alpha}$ are the same as those when the unknown factors are observed. We call this phase the *strong oracle phase.* When $T^{1-2\varphi}$ diverges to infinity at a rate faster than $T^{(\kappa+1-2\varphi)/3}$, the resulting convergence rate and asymptotic distribution for $\widehat{\gamma}$ are different from those under the strong oracle phase. Even in this case, the convergence rate and asymptotic distribution for $\widehat{\alpha}$ are still the same as those when the unknown factors are observed. This corresponds to *weak oracle phase.* The phase transition occurs when $T^{1-2\varphi} = T^{(\kappa+1-2\varphi)/3}$, which is the *semi-strong oracle case* and the

Figure 1: An Example of $A(\omega, g)$



critical boundary of the phase transition. Changes in the convergence rates and asymptotic distributions are continuous along the critical boundary.

Figure 2: Phase Diagram



Notes. This figure depicts a phase transition on the (κ, φ)-space. The possible region we consider on the (κ, φ)-space is $0 < \varphi < 1/2$ and $\kappa \geq 1$. The critical boundary, i.e., the semi-strong oracle region ($\varphi = -\kappa/4 + 1/2$) is shown by closely dotted points in the figure. The strong oracle phase is shaded in blue, wheres the weak oracle phase is in green.

Figure 2 depicts a phase transition from the strong oracle phase to the weak oracle phase. The possible region we consider on the (κ, φ)-space is $0 < \varphi < 1/2$ and $\kappa \geq 1$. The critical boundary ($\varphi = -\kappa/4 + 1/2$) is shown by closely dotted points in the figure. The strong oracle phase is shaded in blue, wheres the weak oracle phase is in green. On the one hand, as φ moves from 0 to 1/2, the strong oracle region for κ increases. That is, as the convergence rate for $\widehat{\gamma}$ gets slower, the requirement for the minimal sample size N for factor estimation becomes less stringent. On the other hand, as κ gets larger, the strong oracle region for φ increases. In other words, as N gets larger, the range of attainable oracle rates of convergence for $\widehat{\gamma}$ becomes wider. In this way, we provide a thorough characterization of the effect of estimated factors.

7 Inference

In this section, we consider inference. Regarding α_0, Theorem 4.1 implies that inference for α_0 can be carried out as if γ_0 were known, provided that γ_0 is identified. The same conclusion holds with estimated f_t, as shown in Theorem 6.3, provided that $T = O(N)$. In sum, standard asymptotic normal inference for α_0 can be carried out for both observed and estimated f_t.

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).

When f_t is known, we propose to use the following statistic:

$$\begin{aligned} \sup Q &= \sup_{\gamma \in \Gamma} Q_T(\gamma) = \sup_{\gamma \in \Gamma} T \frac{\min_{\alpha:\delta=0} \mathbb{S}_T(\alpha, \gamma) - \min_{\alpha} \mathbb{S}_T(\alpha, \gamma)}{\min_{\alpha} \mathbb{S}_T(\alpha, \gamma)} \\ &= T \frac{\min_{\alpha:\delta=0} \mathbb{S}_T(\alpha, \gamma) - \min_{\alpha,\gamma} \mathbb{S}_T(\alpha, \gamma)}{\min_{\alpha,\gamma} \mathbb{S}_T(\alpha, \gamma)}, \end{aligned} \tag{7.1}$$

where $\mathbb{S}_T(\alpha, \gamma)$ is defined in (3.3). Note that $Q_T(\gamma)$ is the likelihood ratio (LR) statistic for $\delta = 0$ when γ is given and the regression error is Gaussian. When f_t is estimated, we suggest adopting the sup$Q$ statistic by replacing $\mathbb{S}_T(\alpha, \gamma)$ with $\widetilde{\mathbb{S}}_T(\alpha, \gamma)$ in (6.4).

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

Theorem 7.1. *If the factor f_t is known, let Assumptions 1, 3, and 4 hold. If the factor f_t is estimated, let Assumptions 1, 5, 6-9 hold and $T = O(N)$. Then, under $\mathcal{H}_0$,*

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

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.[13]

Under conditional heteroskedasticity, the limiting null distribution is different from that of Hansen (1996). His asymptotic null distribution is the supremum of a chi-square process, whereas ours is the supremum of a "weighted" chi-square process. To obtain the former, it is necessary to use a proper scaling matrix to account for the heteroskedastic errors. However, it would be challenging to compute the supremum of the test statistic with a scaling matrix. Furthermore, its asymptotic distribution is still not pivotal, thereby requiring the bootstrap

[13] 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.

to obtain p-values. It is worth noting that our statistic has an advantage in terms of computation because we can utilize the computation algorithm developed in Section 3.1. The computational affordability becomes more important when we employ the bootstrap below.

`[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 (7.1) 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
$$p^* = 1 - F_T^*(\text{supQ}),$$
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 E.1, the maximal inequality in Lemma H.1 and the conditional martingale difference sequence (mds) 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.

It is non-standard to carry out inference on γ because of the non-conventional convergence rates (the estimated factor case, in particular) with the unknown φ and the need to simulate the Gaussian process with a general form of the covariance kernel. We leave it as an interesting topic for future research.

8 Monte Carlo Experiments

In this section we study the finite sample properties of the proposed method via Monte Carlo experiments. The data are generated from the following design:

$$y_t = x_t'\beta_0 + x_t'\delta_0 1\{g_t'\phi_0 > 0\} + \varepsilon_t \quad \text{for} \quad t = 1, \ldots, T,$$

where $\varepsilon_t \sim N(0, 0.5^2)$, $x_t \equiv (1, x_{2,t}')'$, and $g_t \equiv (g_{1,t}', -1)'$. Both x_{2t} and $g_{1,t}$ follow the vector autoregressive model of order 1:

$$x_{2,t} = \rho_x x_{2,t-1} + \nu_t$$
$$g_{1,t} = \rho_g g_{1,t-1} + u_t,$$

where $\nu_t \sim N(0, I_{d_x-1})$ and $u_t \sim N(0, I_K)$. When the factor g_t is not observable, we instead observe $\mathcal{Y}_t$ that is generated from

$$\mathcal{Y}_t = \Lambda g_{1,t} + \sqrt{K} e_t$$
$$e_t = \rho_e e_{t-1} + \omega_t,$$

where $\mathcal{Y}_t$ is an $N \times 1$ vector and ω_t is an i.i.d. innovation generated from $N(0, I_N)$. The terms $\varepsilon_t, \nu_t, u_t, \omega_t$ are mutually independent of each other.

In the baseline model, we use the joint estimation algorithm and consider the case of $T = N = 200$, $d_x = 2$ and $K = 3$. The parameter values are set as follows: $\beta_0 = \delta_0 = (1, 1)$, $\phi_0 = (1, 2/3, 0, 2/3)$, $\rho_x = \text{diag}(0.5, \ldots, 0.5)$, $\rho_g = \text{diag}(\rho_{g,1}, \ldots, \rho_{g,K})$, where $\rho_{g,k} \sim U(0.2, 0.8)$ for $k = 1, \ldots, K$, the i-th row of Λ, $\lambda_i' \sim N(0', K \cdot I_K)$, and $\rho_e = \text{diag}(\rho_{e,1}, \ldots, \rho_{e,N})$, where $\rho_{e,i} \sim U(0.3, 0.5)$ for $i = 1, \ldots, N$. The values of ρ_g and ρ_e are drawn only once and kept for the whole replications. The factor model design is similar to Bai and Ng (2009) and Cheng and Hansen (2015). All simulation results are based on 1,000 replications and are performed on a desktop computer equipped with an AMD RYZEN Threadripper 1950X CPU.

Table 1 summarizes the simulation results of the baseline model. We estimate the model under four different scenarios: (i) when we know the correct regime (Oracle), i.e. ϕ_0 is known; (ii) when we observe g_t and know that the third factor is irrelevant (Observed Factors/No Selection g_t); (iii) when we observe g_t and have to select the relevant factors (Observed Factors/Selection on g_t); and (iv) when we do not observe g_t but estimate factors from $\mathcal{Y}_t$ by the principal component analysis. In the last case, we set the number of feasible factors to be 4. We report the mean bias and the root-mean-square error (RMSE) for β, δ, or γ as well as the coverage rate for the 95% confidence intervals of β and δ. We also report the ratio of samples that the correct factors are selected (Correct Factor Selection) in scenario (iii). In scenarios (ii)–(iv), we report the average of correct regime prediction (Ave. Cor. Regime Prediction). This statistic 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)):

$$\widehat{E}\left(\frac{1}{T}\sum_{t=1}^{T} 1\left\{1\{g_t'\hat{\phi} > 0\} = 1\{g_t'\phi_0 > 0\}\right\}\right),$$

Table 1: Simulation Results: Baseline Model

	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)

where the expectation $\widehat{E}$ is taken over simulation draws. The standard errors are reported in the parentheses next to the statistic. The regime classification results are almost perfect in scenarios (ii) and (iii) and slightly worse in scenario (iv).

Overall, the finite sample performance of the proposed method is satisfactory. As predicted by asymptotic theory, the estimation results of $\alpha = (\beta, \delta)$ in (ii)–(iv) are quite similar to those of the oracle model in (i). The coverage rates for the 95% confidence intervals are also close to the nominal value. Not surprisingly, these results on α are based on the good performance in estimating ϕ (or γ). The method also shows good performance in selecting factors in (iii).

Table 2: 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)	

Table 3: Computation Time for Different Sample Sizes (unit=second)

		T=200	T=300	T=400	T=500
Min	Iter. ($\zeta = 1.0$)	1.46	2.19	2.86	3.68
	Iter. ($\zeta = 0.1$)	1.50	2.24	2.92	3.74
	Joint	1.87	2.85	3.97	5.23
Median	Iter. ($\zeta = 1.0$)	1.49	2.23	2.99	3.78
	Iter. ($\zeta = 0.1$)	1.52	2.27	3.04	3.81
	Joint	1.99	3.04	4.39	5.66
Mean	Iter. ($\zeta = 1.0$)	1.49	2.24	2.99	3.78
	Iter. ($\zeta = 0.1$)	1.54	2.28	3.05	3.83
	Joint	1.99	3.09	4.34	5.66
Max	Iter. ($\zeta = 1.0$)	1.53	2.33	3.09	3.94
	Iter. ($\zeta = 0.1$)	2.54	2.42	3.16	3.98
	Joint	2.21	3.69	4.73	6.07
Convergence Ratio	($\zeta = 1.0$)	0.93	0.87	0.93	0.88
	($\zeta = 0.1$)	1.00	0.97	1.00	0.99

Note: The unit of computation time is second. The convergence ratio measures the proportion that the difference of two objective function values is less than 10^{-6}.

Table 4: Computation Time for Different Sizes of x_t (unit=second)

		$d_x = 1$	$d_x = 2$	$d_x = 3$	$d_x = 4$
Min	Iter. ($\zeta = 1.0$)	1.46	1.45	1.45	1.45
	Iter. ($\zeta = 0.1$)	1.50	1.49	1.49	1.49
	Joint	1.87	2.16	2.39	2.46
Median	Iter. ($\zeta = 1.0$)	1.49	1.49	1.48	1.48
	Iter. ($\zeta = 0.1$)	1.52	1.52	1.51	1.52
	Joint	1.99	2.31	2.52	2.76
Mean	Iter. ($\zeta = 1.0$)	1.49	1.49	1.48	1.48
	Iter. ($\zeta = 0.1$)	1.54	1.52	1.52	1.52
	Joint	1.99	2.30	2.51	2.76
Max	Iter. ($\zeta = 1.0$)	1.53	1.59	1.53	1.57
	Iter. ($\zeta = 0.1$)	2.54	1.68	1.69	1.68
	Joint	2.21	2.54	2.85	3.06
Convergence Ratio	($\zeta = 1.0$)	0.93	0.84	0.87	0.87
	($\zeta = 0.1$)	1.00	0.98	0.94	0.94

Note: The unit of computation time is second. The convergence ratio measures the proportion that the difference of two objective function values is less than 10^{-6}.

Table 5: Computation Time for Different Sizes of g_t (unit=second)

		$d_g = 2$	$d_g = 3$	$d_g = 4$	$d_g = 5$
Min	Iter.($\zeta = 1.0$)	1.46	1.50	1.83	3.30
	Joint	1.87	2.04	4.79	78.78
Median	Iter.($\zeta = 1.0$)	1.49	1.57	1.92	3.41
	Joint	1.99	2.17	6.42	410.35
Mean	Iter.($\zeta = 1.0$)	1.49	1.57	1.93	3.43
	Joint	1.99	2.18	6.56	445.15
Max	Iter.($\zeta = 1.0$)	1.53	1.66	2.26	3.68
	Joint	2.21	2.38	9.68	1389.86
Convergence Ratio	($\zeta = 1.0$)	0.93	0.94	0.88	0.92

Note: The unit of computation time is second. The convergence ratio measures the proportion that the difference of two objective function values is less than 10^{-6}.

In Table 2 we focus on the unobserved factor model and check the performance of the estimator by increasing N. For each simulated sample of $\{y_t, x_t, g_t\}$, we generate $\mathcal{Y}_t$ with $N = 100, 200, 400, 1600$. We use the same baseline design with $T = 200$, $d_x = 2$, but $K = 1$. We have chosen the simpler specification $K = 1$ to speed up computations in this experiment. We use the joint estimation algorithm and conduct 1,000 replications. The regimes are predicted more precisely as N increases and the performance of the estimator improves. We observe relatively more improvements in γ rather than α. This is because $\widehat{\alpha}$ enjoys the oracle property, provided that $T = O(N)$.

Finally, Tables 3–5 report summary statistics of computation time as well as the convergence ratio of each computation method. Specifically, the convergence ratio measures the proportion such that the difference of two objective function values is less than 10^{-6}. We simplify the baseline model by considering only observed factors and by setting $\rho_x = \rho_g = 0$, i.e. no serial dependency in x_t and g_t. The results are based on 100 replications. We consider scenario (ii), so the correct factors are observed and we do not need to select them. We set $T = 200$, $d_x = 1$, and $d_g = 2$, initially and increase each dimension as follows.

First, we vary the sample size $T = \{200, 300, 400, 500\}$. For the iterative method, we consider a coarse grid ($\zeta = 1.0$) and a fine grid ($\zeta = 0.1$). Recall that ζ is the minimum distance between two grid points. Thus, given the lower and upper bounds of γ_j, $\underline{\gamma}_j$ and $\overline{\gamma}_j$, we set the grid points as $\{(1, \gamma_2, \ldots, \gamma_{d_f}) : \underline{\gamma}_j + (k-1)\zeta$ for all integer k such that $1 \leq k \leq 1 + \zeta^{-1}(\overline{\gamma}_j - \underline{\gamma}_j)$ and $j = 2, \ldots, d_f\}$. In total, there are $\Pi_{j=2}^{d_f}[1 + \zeta^{-1}(\overline{\gamma}_j - \underline{\gamma}_j)]$ grid points. In Table 3, the computation time of all methods increases as T increases but all of them deliver the computation results in a reasonable range of time (about 6 seconds in the worse case). The iteration method with a coarse grid is the fastest but it sometimes ends up with local minima (13% of simulations in the worst case). Table 4 summarizes the result when we increase the dimension of x_t, $d_x = \{1, 2, 3, 4\}$ while keeping $T = 200$ and $d_g = 2$. Both iterative methods do not lose the computation time while the joint method gets slower as d_x increases. However, there is a trade-off between the fast computation and the convergence rate. Even with a fine grid ($\zeta = 0.1$), about 6% of the simulations end up with some local minima. In Table 5, we increase $d_g = \{2, 3, 4, 5\}$ while keeping $T = 200$ and $d_x = 1$. The grid search in the iterative method with $\zeta = 0.1$ takes longer than a reasonable range of computation time and we only report the result with $\zeta = 1.0$. As d_g increases, computation time required for the joint method increases exponentially but still stays in the feasible range. The iterative method is faster but it finds local minima around 10% of simulations. Therefore, if one has a model with a large dimension of g_t or f_t, we recommend estimating it first by the iterative method with a coarse grid but producing the final result by the joint method.

9 Empirical Examples

9.1 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:

$$
\begin{aligned}
y_t &= \underset{(2.12)}{-3.21} + \underset{(0.25)}{0.51} y_{t-1} - \underset{(0.31)}{0.93} y_{t-2} - \underset{(0.25)}{0.38} y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} \leq 0.01 \\
y_t &= \underset{(0.77)}{2.14} + \underset{(0.10)}{0.30} y_{t-1} + \underset{(0.10)}{0.18} y_{t-2} - \underset{(0.07)}{0.16} y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} > 0.01,
\end{aligned}
\tag{9.1}
$$

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 (9.1). 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:

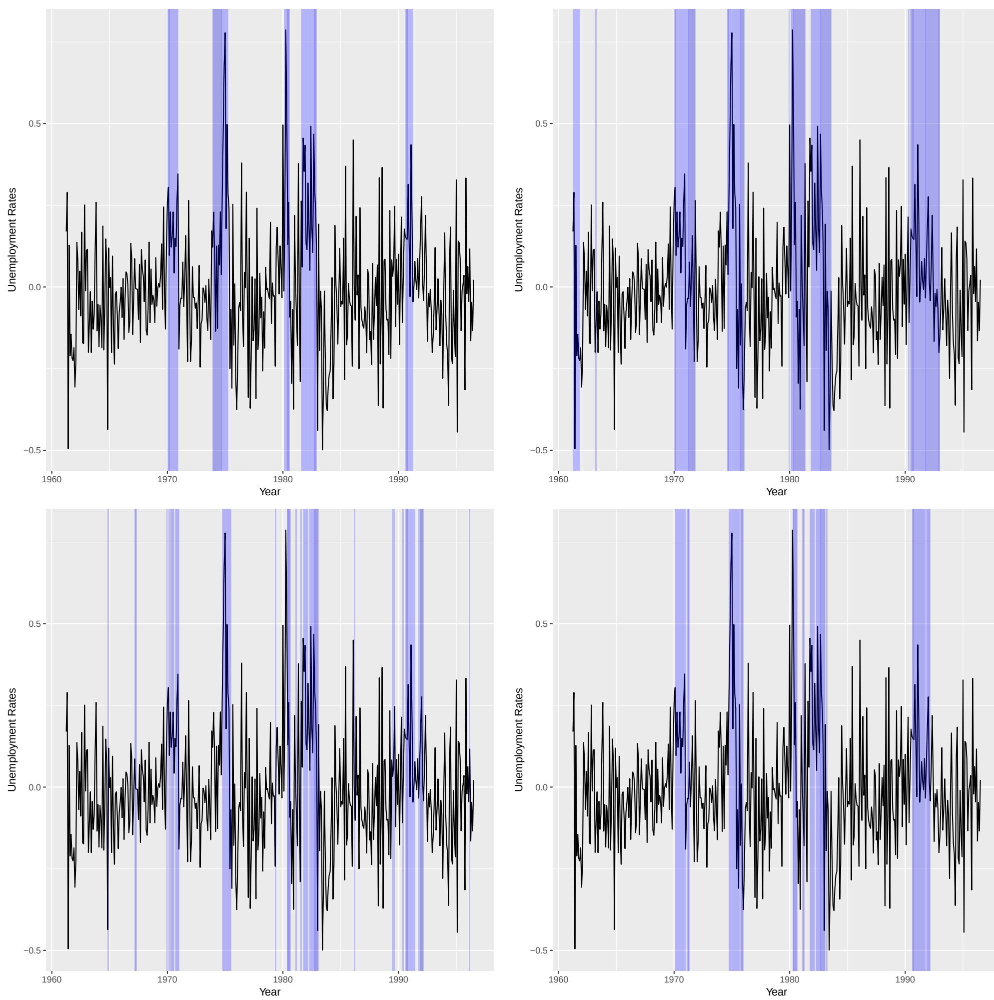
$$\begin{aligned}
y_t &= \underset{(1.33)}{-2.07} + \underset{(0.13)}{0.28}y_{t-1} - \underset{(0.16)}{0.33}y_{t-2} + \underset{(0.19)}{0.62}y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} - 0.91y_{t-5} \le -0.50 \\
y_t &= \underset{(0.96)}{2.76} + \underset{(0.12)}{0.35}y_{t-1} + \underset{(0.12)}{0.07}y_{t-2} - \underset{(0.10)}{0.21}y_{t-5} + \widehat{\varepsilon}_t \quad \text{if } y_{t-2} - 0.91y_{t-5} > -0.50.
\end{aligned} \tag{9.2}$$

We now report the result of testing the null hypothesis of no threshold effect. We take our estimates in (9.2) 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}.

9.2 Classifying the Regimes of US Unemployment

Following Hansen (1997), we now consider threshold autoregressive models for the US unemployment rate. Hansen (1997) used monthly unemployment rates for males age 20 and over and estimated his threshold model with the first-differenced series, say Δy_t, to avoid nonstationarity. The leg length in the autoregressive model was $p = 12$ and his preferred threshold variable was $q_{t-1} = y_{t-1} - y_{t-12}$. In this section, we investigate the usefulness of using unknown but estimated factors. We use the first factor, say F_t, of Ludvigson and Ng (2009) among eight common factors that are estimated from 132 macroeconomic variables. This factor not only explains the largest fraction of the total variation in their panel data set but also loads heavily on employment, production, and so on. They call it a *real factor* and thus it is a legitimate candidate for explaining the unemployment rate. We consider three different specifications for f_t: (1) $f_{1t} = (q_{t-1}, -1)$, (2) $f_{2t} = (F_{t-1}, -1)$, and (3) $f_{3t} = (q_{t-1}, F_{t-1}, -1)$. That is, the first specification of f_t corresponds to Hansen (1997), the second one uses the real factor only, and the third case includes both. We combined the updated estimates of the real factor, which are available on Ludvigson's web page, with Hansen's data, yielding a monthly sample from March 1960 to July 1996 for our estimation purpose.

Figure 3: Regime Classification



Note. The top left panel shows NBER recession dates in the shaded area, the top right panel displays regime 1 with specification (1), and the bottom left and right panels show regime 1 with specifications (2) and (3), respectively.

Table 6 reports the parameter estimates of regression coefficients and their heteroskedasticity consistent standard errors for each of three specifications. The estimated intercept is negative in regime 1 but positive in regime 2 across all three specifications. Hence, we label

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

regime 1 "contraction" and regime 2 "expansion", respectively. Point estimates of lagged unemployment rates indicate different dynamics across different specifications; however, it might be more illuminating to consider the overall performance of different models. For this purpose, in Table 6, we show the goodness of fit by reporting the average of squared residuals and also the results of regime classification relative to the NBER business cycle dates. The latter is obtained by

$$1 - \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{f_{jt}'\widehat{\gamma}_j > 0\right\} - 1_{\mathrm{NBER},t}\right| \quad \text{for each } j = 1, 2, 3,$$

where $\widehat{\gamma}_j$ is the parameter estimate when factor f_{jt} is considered and $1_{\mathrm{NBER},t}$ is the indicator function that has value 1 if and only if the economy is in expansion according to the NBER dates. Figure 3 gives the graphical representation of regime classification. Specification (1) suffers from the highest level of mis-classification and tends to classify recessions more often than the NBER; specification (2) mitigates the misclassification risk but at the expense of a worse goodness of fit. On one hand, the threshold autoregressive model solely by q_{t-1} fittingly explains the unemployment rate but is short of classifying the overall economic conditions satisfactorily; on the other hand, the model based only on F_{t-1} is adequate at describing the underlying overall economy but is not reaching as far as the former model in terms of explaining the unemployment rate. It turns out that specification (3) enjoys advantages of both specifications (1) and (2). It has the lowest misclassification error and best explains unemployment. Thus, we have shown the real benefits of using a vector of possibly unobserved factors to explain the unemployment dynamics.

As an additional check, we tested the null hypothesis of no threshold effect. We take our estimates in specification (3) as unconstrained estimates. The resulting p-value is 0.002 based on 500 bootstrap replications, thus providing strong evidence for the existence of two regimes.

10 Conclusions

We have proposed a new method for estimating a two-regime regression model where the switching between the regimes is driven by a vector of possibly unobservable factors. We have shown that our optimization problem can be reformulated as mixed integer optimization and have presented two alternative computational algorithms. We have also derived the asymptotic distribution of the resulting estimator under the scheme that the threshold effect shrinks to zero as the sample size tends to infinity. We have demonstrated that our proposed method works well in finite samples and have illustrated its usefulness by applying it to US macro data.

There are several areas that this paper did not cover. First, it might be fruitful to build on the literature on threshold models with endogeneity (see, e.g., Caner and Hansen, 2004; Seo and Shin, 2016; Yu and Phillips, 2018) and extend our framework in that direction. Second, we might consider a setup of high dimensional regression models as in Lee, Seo, and Shin (2016) and Lee, Liao, Seo, and Shin (2018) and allow for regime classification by high dimensional factors. Third, as an alternative measure of factors, one may consider an index of economic policy uncertainty based on newspaper coverage frequency (Baker, Bloom, and Davis, 2016) or measures of the conditional volatility of an unforecastable disturbance, constructed from macroeconomic and financial indicators or from firm-level microdata (Jurado, Ludvigson, and Ng, 2015). More recently, Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2018) developed empirical measures of uncertainty using detailed Census microdata and concluded that recessions are best modelled as being driven by shocks with a negative first moment and a positive second moment. This suggests that we could include both first and second moment shocks as factors. These are possible directions for future research.

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.

Auerbach, A. J., and Y. Gorodnichenko (2012): "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4(2), 1–27.

——— (2013a): "Fiscal Multipliers in Recession and Expansion," in *Fiscal Policy after the Financial Crisis*, ed. by A. Alesina, and F. Giavazzi, pp. 63–98. University of Chicago Press.

——— (2013b): "Output Spillovers from Fiscal Policy," *American Economic Review*, 103(3), 141–46.

Bai, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

Bai, J., and S. Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133–1150.

——— (2008): "Extremum Estimation when the Predictors are Estimated from Large Panels," *Annals of Economics and Fiance*, 9(2), 201–222.

——— (2009): "Boosting diffusion indices," *Journal of Applied Econometrics*, 24(4), 607–629.

Bai, J., and P. Wang (2016): "Econometric Analysis of Large Factor Models," *Annual Review of Economics*, 8(1), 53–80.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *Quarterly Journal of Economics*, 131(4), 1593–1636.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120(1), 387–422.

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.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really Uncertain Business Cycles," *Econometrica*, 86(3), 1031–1065.

Caner, M., and B. E. Hansen (2001): "Threshold Autoregression with a Unit Root," *Econometrica*, 69(6), 1555–1596.

——— (2004): "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20(5), 813–843.

Chan, K.-S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *Annals of Statistics*, 21(1), 520–533.

Cheng, X., and B. E. Hansen (2015): "Forecasting with factor-augmented regression: A frequentist model averaging approach," *Journal of Econometrics*, 186(2), 280–293.

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.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The generalized dynamic factor model: identification and estimation," *Review of Economics and Statistics*, 82, 540–554.

Galvão, A. B., and T. Owyang, Michael (2018): "Financial Stress Regimes and the Macroeconomy," *Journal of Money, Credit and Banking*, 50(7), 1479–1505.

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.

——— (1997): “Inference in TAR Models,” *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1–14.

——— (1999): “Threshold effects in non-dynamic panels: Estimation, testing, and inference,” *Journal of Econometrics*, 93(2), 345–368.

——— (2000): “Sample splitting and threshold estimation,” *Econometrica*, 68(3), 575–603.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): “Measuring Uncertainty,” *American Economic Review*, 105(3), 1177–1216.

Kim, J., and D. Pollard (1990): “Cube Root Asymptotics,” *Annals of Statistics*, 18(1), 191–219.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2018): “Oracle Estimation of a Change Point in High-Dimensional Quantile Regression,” *Journal of the American Statistical Association*, 113(523), 1184–1194.

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.

——— (2016): “The lasso for high dimensional regression with a possible change point,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1), 193–210.

Ludvigson, S. C., and S. Ng (2009): “Macro Factors in Bond Risk Premia,” *Review of Financial Studies*, 22(12), 5027–5067.

Merlevède, F., M. Peligrad, and E. Rio (2011): “A Bernstein type inequality and moderate deviations for weakly dependent sequences,” *Probability Theory and Related Fields*, 151(3), 435–474.

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,” *Journal of Applied Econometrics*, 10(2), 109–125.

Ramey, V., and S. Zubairy (2018): “Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data,” *Journal of Political Economy*, 126(2), 850–901.

Seijo, E., and B. Sen (2011): “Change-point in stochastic design regression and the bootstrap,” *Annals of Statistics*, 39(3), 1580–1607.

Seo, M. H., and O. Linton (2007): “A smoothed least squares estimator for threshold regression models,” *Journal of Econometrics*, 141(2), 704–735.

Seo, M. H., and Y. Shin (2016): “Dynamic panels with threshold effect and endogeneity,” *Journal of Econometrics*, 195(2), 169–186.

Stock, J., and M. Watson (2002a): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Stock, J. H., and M. W. Watson (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20(2), 147–162.

Tenreyro, S., and G. Thwaites (2016): "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions," *American Economic Journal: Macroeconomics*, 8(4), 43–74.

Tong, H. (1990): *Non-linear time series: a dynamical system approach*. Oxford University Press.

van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes*. Springer, New York.

Wang, Y. (2018): "Inference in the Change Point Model with Time Varying Moments," Available at `https://sites.google.com/site/yulongwanghome/`.

Yu, P., and P. C. Phillips (2018): "Threshold regression with endogeneity," *Journal of Econometrics*, 203(1), 50–68.

Online Appendices to "Factor-Driven Two-Regime Regression" by Lee, Liao, Seo and Shin

A Proof of Identification in Section 2

Proof of Theorem 2.1. 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 $B_\gamma = \{f_t'\gamma_0 \le 0 < f_t'\gamma\} \cup \{f_t'\gamma \le 0 < f_t'\gamma_0\}$. Thus,

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

by (2.2) and $R(\alpha_0,\gamma)$ is continuous at $\gamma = \gamma_0$ due to Assumption 2 (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 $\{f_t'\gamma_0 \le 0\} \cap \{f_t'\gamma \le 0\}$. Thus,

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

for some $c > 0$ due to the rank condition in (2.3).

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 3.1, present an alternative form of the proposed algorithm in Section 3.1, describe additional possible restrictions in estimation and give practical guidance.

B.1 Proof for Section 3

Proof of Theorem 3.1. 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 $\widehat{d}_t = 1\{f_t'\widehat{\gamma} > 0\}$. 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 Joint Optimization

The proposed algorithm in Section 3.1 may run slowly when the dimension of x_t is large. To mitigate this problem, we reformulate the joint optimization in the following way.

`[Joint Optimization (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

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

for each $t=1,\dots,T$ and each $j=1,\dots,d_x$, 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 $j=1,\dots,d_x$, while ensuring that optimization problem (B.1) under (B.2) is mathematically equivalent to optimization problem (3.5) under (3.6) in Section 3.1. We use the alternative form of formulation in our numerical work; however, we present a simpler form in Section 3.1 to help readers follow our basic ideas more easily.

B.3 Additional restrictions

We may also consider

$$
\frac{1}{T}\sum_{t=2}^{T}|d_{t+1}-d_t|\le M \tag{B.3}
$$

for some predetermined $M>0$. This restriction limits the maximum number of regime changes. To impose (B.3) in mixed integer programming, introduce $\Delta_{t+1},\Delta_{t+1}^{+},\Delta_{t+1}^{-}$ such that

$$
\begin{aligned}
&\Delta_{t+1} = d_{t+1} - d_t, \\
&\Delta_{t+1} = \Delta_{t+1}^{+} - \Delta_{t+1}^{-}, \\
&(\Delta_{t+1}^{+}, \Delta_{t+1}^{-}) : \ \text{SOS-1}, \\
&\frac{1}{T}\sum_{t=2}^{T}\left[\Delta_{t+1}^{+} + \Delta_{t+1}^{-}\right] \leq M, \\
&\Delta_{t+1}^{+} \in \{0,1\}, \\
&\Delta_{t+1}^{-} \in \{0,1\}
\end{aligned}
$$

for each $t = 2, \ldots, T$. Here, $(\Delta_{t+1}^{+}, \Delta_{t+1}^{-}) :$ SOS-1 refers to Specially Ordered Sets of type 1, which means that at most one of Δ_{t+1}^{+} and Δ_{t+1}^{-} may take a non-zero value.

Alternatively,

$$
\frac{1}{T}\sum_{t=k+1}^{k+m} |d_{t+1} - d_t| \leq 1 \ \text{ for each } k \leq T - m \tag{B.4}
$$

for some predetermined $m > 0$. This imposes that only one change is allowed within the m time periods. The restriction (B.4) can also be written as the SOS-1 type constraint.

B.4 Practical Guidance

We have presented two alternative classes of MIO algorithms. The first one is a global approach that ensures that its solution is globally optimal once it is found. The second one is an iterative approach that typically computes much faster in problems with a much large T. Though it does not guarantee that the resulting solution is globally optimal, it produces an asymptotically equivalent estimator of $(\alpha_0', \gamma_0')'$. In addition, we find that it works pretty well in our applications even when the size m_T of Γ_T is relatively small and the number of iterations in Steps 3(a)-(c) is less than three.

As such, we view that both are complements to each other. On one hand, when T is relatively small, we recommend using the first approach; on the other hand, when T is relatively large or we need to estimate parameters repeatedly, we advise practitioners to use the second approach. In practice, one may combine both methods. For example, one could use the iterative approach to obtain an initial estimator and switch to the joint approach to obtain a final estimator in a narrowly defined parameter space around the initial estimator.

C Proofs of the Asymptotic Distribution in Section 4: Known f

Recall that we have proposed two (asymptotically equivalent) estimators for (α, γ). One is defined as the global minimizer of the least squares problem, jointly solved by applying the MIQP. The other is defined by iteratively solving the MIO problem using MILP. We shall show that both estimators have the same asymptotic distribution. We split the proofs into two parts: the case of the joint approach and that of the iterative approach.

C.1 Case 1: Joint Approach

We start with the joint approach. The proof is divided into the following subsections.

C.1.1 Consistency

Lemma C.1 (Consistency)**.** *Let Assumptions 1, 2 and 3 (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 C.1. We begin with stating the following standard ULLN for ρ-mixing sequences, see e.g. Davidson (1994), for which Assumption 3 (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

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

Then, by Lemma C.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 (iii), the fact that $P(\gamma_0)X_0 = X_0$,

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

However, the term $A(\widetilde{\gamma})$ is continuous by Assumption 2 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 2 (ii) and 3 (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

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

where

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

First, note that

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

by ULLN. Similarly,

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

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 2.1 implies that $\widehat{\alpha}-\alpha_0=o_P(1)$ as (A.1) shows that R is bounded below by some positive constant times $|\alpha-\alpha_0|_2^2$. ∎

C.1.2 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

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

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 H.1 and H.2 for an easy reference.

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

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

where $k=1,2,3$.

Lemma C.3 (Rates of Convergence)**.** *Let Assumptions 1, 2, 3, and 4 hold. Then as* $T \to \infty$*,*

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

Proof of Lemma C.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,

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

for any α and γ such that $|\alpha - \alpha_0| < e$ and $|\gamma - \gamma_0| < e$. Recall $R(\alpha, \gamma)$ is defined in (2.1).

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

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

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

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

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

we conclude that

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

That is,

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

implying

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

∎

Proof of Step 1. Due to Assumption 4 and then Assumption 2 we can find positive constants c, c_0 such that

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

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

$$G(\gamma) = \mathbb{E}\left|1_t(\gamma) - 1_t(\gamma_0)\right| \geq c\left|\gamma - \gamma_0\right|_2.$$

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

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

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,

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

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,

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

Thus,

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

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 4.

Next,

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

due to Assumption 3 (iii).

Also, note that

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

where the second inequality comes from Assumption 3 (i) and Assumption 2 (i). Combining the inequalities above together yields that

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

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,

$$
R(\alpha,\gamma) \geq \frac{c_1}{2}\left|\alpha-\alpha_0\right|_2^2 + c_0T^{-2\varphi}\left|\gamma-\gamma_0\right|_2.
$$

When (ii) holds, we have that

$$
C_2T^{-\varphi}\left|\gamma-\gamma_0\right|_2\left|\alpha-\alpha_0\right|_2 < 2c_1^{-1}C_2^2T^{-2\varphi}\left|\gamma-\gamma_0\right|_2^2.
$$

Then under (ii),

$$
\begin{aligned}
&c_0T^{-2\varphi}\left|\gamma-\gamma_0\right|_2 - C_2T^{-\varphi}\left|\gamma-\gamma_0\right|_2\left|\alpha-\alpha_0\right|_2 \\
&> T^{-2\varphi}\left|\gamma-\gamma_0\right|_2\left[c_0 - 2c_1^{-1}C_2^2\left|\gamma-\gamma_0\right|_2\right].
\end{aligned}
$$

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

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

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

for any $0 < \eta < c/2$, by the MDS CLT and Lemma H.1 for the first term $T^{-1/2}\sum_{t=1}^{T}\varepsilon_tZ_t(\gamma)$ and by Assumption C.2 for the second term $T^{-1}\sum_{t=1}^{T}\varepsilon_tx_t'\delta_0\left(1_t(\gamma)-1_t(\gamma_0)\right)$.

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

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

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

C.1.3 Asymptotic Distribution

Proof of Theorem 4.1. 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 (C.10) and Lemma C.2 together imply that

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

Recall (C.8) and write that

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

Then, due to Assumption 4,

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

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

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

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 H.1,

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

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

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

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

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

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

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

Step 2.

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

where W is a Gaussian process whose covariance kernel is given by $H(\cdot,\cdot)$ in (4.1) 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 H.1 since for any finite g and for any $c>0$,

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

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, $T^{-2}\mathbb{E}\max_t w_t^4\le T^{-1}\mathbb{E}w_t^4$ since $\max_t|a_t|\le\sum_{t=1}^{T}|a_t|$ and w_t is stationary. Now,

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

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 C.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,

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

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

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

∎

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 C.1 that addresses the independence between the processes of a and g. ∎

C.2 Case 2: Iterative Approach

The proofs for the iterative approach are similar to those in the previous subsection but with some different details. For the completeness of the proofs, we provide full details for this case as well. In particular, we prove Theorem 4.1 through the following claims.

Claim 1. $\widehat{\gamma}^0 \xrightarrow{p} \gamma_0$ for the approximate estimate $\widehat{\gamma}^0 = \operatorname{argmin}_{\gamma \in \Gamma_T} \mathbb{S}_T(\gamma)$.

Claim 2. For a given γ, let

$$\widehat{\alpha}(\gamma) = \underset{\alpha}{\operatorname{argmin}}\, \mathbb{S}_T(\alpha, \gamma).$$

Then, for any $\vec{\gamma} \xrightarrow{p} \gamma_0$,

$$T^{\varphi}\left(\widehat{\alpha}\left(\vec{\gamma}\right) - \alpha_0\right) = o_P\left(1\right).$$

Claim 3. For a given α, let

$$\widehat{\gamma}\left(\alpha\right) = \underset{\gamma \in \Gamma}{\operatorname{argmin}}\, \mathbb{S}_T\left(\alpha, \gamma\right).$$

Then, for any $\vec{\alpha} = \alpha_0 + o_P\left(T^{-\varphi}\right)$,

$$\widehat{\gamma}\left(\vec{\alpha}\right) - \gamma_0 = O_P\left(T^{-1+2\varphi}\right),$$

and

$$\widehat{\gamma}\left(\vec{\alpha}\right) - \widehat{\gamma}\left(\alpha_0\right) = o_P\left(T^{-1+2\varphi}\right).$$

Claim 4. For $\vec{\gamma} = \gamma_0 + O_P\left(T^{-1+2\varphi}\right)$,

$$\widehat{\alpha}\left(\vec{\gamma}\right) = \widehat{\alpha}\left(\gamma_0\right) + o_P\left(\frac{1}{\sqrt{T}}\right).$$

Claim 5. Derive the asymptotic independence of $T^{1-2\varphi}\left(\widehat{\gamma}\left(\vec{\alpha}\right) - \gamma_0\right)$ and $\sqrt{T}\left(\widehat{\alpha}\left(\vec{\gamma}\right) - \alpha_0\right)$ and their marginal asymptotic distributions.

Then, for our iterative estimates, we can easily note that $\widehat{\alpha}^0 = \widehat{\alpha}\left(\widehat{\gamma}^0\right)$ fulfils the conditions for claim 2 and $\widehat{\gamma}^1$ does for claim 3 as $\widehat{\gamma}^1 = \widehat{\gamma}\left(\widehat{\alpha}^0\right)$, while $\widehat{\alpha}^1$ fits to claim 4 as $\widehat{\alpha}^1 = \widehat{\alpha}\left(\widehat{\gamma}^1\right)$.

Proof of claim 1. It is sufficient to show that $\widehat{\gamma}^0$ satisfies (C.1) in the proof of Lemma C.1. Repeating the argument using Lemma C.2 and the ULLN for the preceding derivation, we can observe that for any $c > 0$ there exists $T_0 < \infty$ such that for all $T > T_0$,

$$\begin{aligned}
& \mathbb{S}_T\left(\widetilde{\gamma}\right) - \mathbb{S}_T\left(\gamma_0\right) \\
= & \min_{\gamma \in \Gamma_T} \mathbb{S}_T\left(\gamma\right) - \mathbb{S}_T\left(\gamma_0\right) \leq \max_{|\gamma - \gamma_0| \leq \psi_T} \left|\mathbb{S}_T\left(\gamma\right) - \mathbb{S}_T\left(\gamma_0\right)\right| \\
= & \frac{1}{T} \max_{|\gamma - \gamma_0| \leq \psi_T} \left|e'\left(P\left(\gamma_0\right) - P\left(\gamma\right)\right) e + 2\delta_0' X_0\left(P\left(\gamma_0\right) - P\left(\gamma\right)\right) e + \delta_0' X_0'\left(P\left(\gamma_0\right) - P\left(\gamma\right)\right) X_0 \delta_0\right| \\
\leq & O_P\left(\frac{1}{\sqrt{T}}\right) + O_P\left(\frac{T^{-\varphi}}{\sqrt{T}}\right) + o_P\left(T^{-2\varphi}\right) + O\left(T^{-2\varphi} c\right) \\
= & o_P\left(T^{-2\varphi}\right),
\end{aligned}$$

where the first inequality is due to the construction of the grid Γ_T and $O\left(T^{-2\varphi} c\right)$ in the last inequality is due to the ULLN for and the continuity of $\operatorname{plim}_{T \to \infty} d_0' X_0' P\left(\gamma\right) X_0 d_0$ at $\gamma = \gamma_0$ due to Assumption 3 (i), while the last equality follows from the fact that c is arbitrary. ∎

Proof of claim 2. By the ULLN and Lemma C.2

$$\begin{aligned}\widehat{\alpha}(\gamma)-\alpha_0 &= \left(\frac{1}{T}\sum_{t=1}^{T} Z_t(\gamma) Z_t(\gamma)'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t(\gamma)\varepsilon_t+\frac{1}{T}\sum_{t=1}^{T} Z_t(\gamma) x_t'\delta_0\left(1_t(\gamma)-1_t\right)\right)\\ &= O_P(1)\left(O_P\left(\frac{1}{\sqrt{T}}\right)+O_P\left(T^{-2\varphi}\left|\gamma-\gamma_0\right|_2\right)+\mathbb{E}Z_t(\gamma) x_t'\delta_0\left(1_t(\gamma)-1_t\right)\right),\end{aligned} \tag{C.17}$$

where $\mathbb{E}\left|Z_t(\gamma) x_t'\delta_0\left(1_t(\gamma)-1_t\right)\right| \le O\left(T^{-\varphi}\left|\gamma-\gamma_0\right|_2\right)$ by Assumption 2 (i) and 3 (i). Then the result follows by setting $\gamma=\vec{\gamma}\xrightarrow{p}\gamma_0$. ∎

Proof of claim 3. Note hat for $\gamma=\widehat{\gamma}(\vec{\alpha})$

$$\begin{aligned} 0 &\ge \left(\mathbb{S}_T(\vec{\alpha},\gamma)-\mathbb{S}_T(\vec{\alpha},\gamma_0)\right)\\ &= \vec{\delta}'\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\left|1_t(\gamma)-1_t\right|\vec{\delta}-\frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\left(1_t(\gamma)-1_t\right)\vec{\delta}\\ &\quad+\left(\vec{\alpha}-\alpha_0\right)'\frac{2}{T}\sum_{t=1}^{T} Z_t(\gamma_0) x_t'\left(1_t(\gamma)-1_t\right)\vec{\delta}.\end{aligned}$$

Then, by the ULLN and the condition for $\vec{\alpha}$,

$$T^{2\varphi}\left(\mathbb{S}_T(\vec{\alpha},\gamma)-\mathbb{S}_T(\vec{\alpha},\gamma_0)\right)\xrightarrow{p}\mathbb{E}\left(d_0'x_t\right)^2\left|1_t(\gamma)-1_t\right|\ge 0,$$

uniformly over $\gamma\in\Gamma$ and the equality holds only when $\gamma=\gamma_0$ by Assumption 2 (ii). Since the limit is continuous by Assumption 2 (i), the argmax continuous mapping theorem yields the consistency of $\widehat{\gamma}(\vec{\alpha})$.

For $\gamma=\widehat{\gamma}(\vec{\alpha})$ in a neighborhood of γ_0, we show that there is $c>0$ such that

$$\begin{aligned} 0 &\ge \left(\mathbb{S}_T(\vec{\alpha},\gamma)-\mathbb{S}_T(\vec{\alpha},\gamma_0)\right)\\ &= \vec{\delta}'\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\left|1_t(\gamma)-1_t\right|\vec{\delta}-\frac{2}{T}\sum_{t=1}^{T}\varepsilon_t x_t'\left(1_t(\gamma)-1_t\right)\vec{\delta}\\ &\quad+\left(\vec{\alpha}-\alpha_0\right)'\frac{2}{T}\sum_{t=1}^{T} Z_t(\gamma_0) x_t'\left(1_t(\gamma)-1_t\right)\vec{\delta}\\ &\ge O_P\left(\frac{1}{T}\right)+cT^{-2\varphi}\left|\gamma-\gamma_0\right|,\end{aligned} \tag{C.18}$$

where $O_P(\cdot)$ is independent of γ. Specifically, we apply Lemma C.2 to the three terms to get

$$\begin{aligned}\delta_0'\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\left|1_t(\gamma)-1_t\right|\delta_0 &= O_P\left(\frac{1}{T^{1+\varphi}}\right)+\left|\delta_0\right|_2\eta T^{-2\varphi}\left|\gamma-\gamma_0\right|_2+T^{-2\varphi}\mathbb{E}\left(d_0'x_t\right)^2\left|1_t(\gamma)-1_t\right| \\ &\geq O_P\left(\frac{1}{T}\right)+cT^{-2\varphi}\left|\gamma-\gamma_0\right|_2,\end{aligned}$$

where the last inequality follows since η is arbitrary while

$$\begin{aligned}&\mathbb{E}\left(d_0'x_t\right)^2\left|1_t(\gamma)-1_t\right| \\ =\ &\mathbb{E}\left[\mathbb{E}\left[\left(d_0'x_t\right)^2|f_t=\gamma\right]\left(1\left\{f_t\gamma\leq 0<f_t'\gamma_0\right\}+1\left\{f_t\gamma_0\leq 0<f_t'\gamma\right\}\right)\right] \\ \geq\ &C\left|\gamma-\gamma_0\right|_2,\end{aligned}$$

for some $C>0$, due to Assumption 4 and Assumption 3 (i). Similarly, we deduce

$$\frac{2}{T}\sum_{t=1}^{T}\varepsilon_tx_t'\left(1_t(\gamma)-1_t\right)\delta_0=O_P\left(\frac{1}{T}\right)+\eta T^{-2\varphi}\left|\gamma-\gamma_0\right|_2 \tag{C.19}$$

$$\left(\vec{\alpha}-\alpha_0\right)'\frac{2}{T}\sum_{t=1}^{T}Z_t(\gamma_0)x_t'\left(1_t(\gamma)-1_t\right)\delta_0=o_P\left(T^{-\varphi}\right)\left(O_P\left(\frac{1}{T}\right)+\eta T^{-2\varphi}\left|\gamma-\gamma_0\right|_2+T^{-\varphi}\left|\gamma-\gamma_0\right|_2\right), \tag{C.20}$$

where η can be arbitrarily chosen. Therefore, combining these results with $\vec{\delta}=\delta_0+o_P\left(T^{-\varphi}\right)$ yields the desired lower bound in (C.18) and thus $\widehat{\gamma}(\vec{\alpha})=\gamma_0+O_P\left(T^{-1+2\varphi}\right)$. Furthermore, (C.19) and (C.20) imply that for any $K<\infty$,

$$\begin{aligned}&\sup_{|\gamma-\gamma_0|\leq KT^{-1+2\varphi}}\left|\mathbb{S}_T(\vec{\alpha},\gamma)-\mathbb{S}_T(\vec{\alpha},\gamma_0)-\left(\mathbb{S}_T(\alpha_0,\gamma)-\mathbb{S}_T(\alpha_0,\gamma_0)\right)\right| \\ &\leq 2\left|\left(\vec{\delta}-\delta_0\right)'\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\left|1_t(\gamma)-1_t\right|\delta_0\right| \\ &+\left|\left(\vec{\delta}-\delta_0\right)'\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\left|1_t(\gamma)-1_t\right|\left(\vec{\delta}-\delta_0\right)\right|+o_P\left(T^{-1}\right) \\ &=o_P\left(T^{-1}\right),\end{aligned} \tag{C.21}$$

by reiterating the argument for (C.20). However, Section C.1.3 shows that $T^{1-2\varphi}\left(\widehat{\gamma}(\vec{\alpha})-\gamma_0\right)$ and $T^{1-2\varphi}\left(\widehat{\gamma}(\alpha_0)-\gamma_0\right)$ are asymptotically equivalent to the argmin of the weak limit of $T\left(\mathbb{S}_T\left(\vec{\alpha},\gamma_0+g\cdot T^{-1+2\varphi}\right)-\mathbb{S}_T(\vec{\alpha},\gamma_0)\right)$ and that of $T\left(\mathbb{S}_T\left(\alpha_0,\gamma_0+g\cdot T^{-1+2\varphi}\right)-\mathbb{S}_T(\alpha_0,\gamma_0)\right)$, respectively. Therefore, the difference between the two processes are $o_P(1)$ due to (C.21), implying that $\widehat{\gamma}(\vec{\alpha})=\widehat{\gamma}(\alpha_0)+o_P\left(T^{-1+2\varphi}\right)$. ∎

Proof of claim 4. From (C.17) in the proof of claim 2, it is sufficient to show that

$$\text{(i)} \quad \left(\frac{1}{T} \sum_{t=1}^{T} Z_t(\vec{\gamma}) Z_t(\vec{\gamma})' \right)^{-1} \xrightarrow{p} \left(\mathbb{E} Z_t(\gamma_0) Z_t(\gamma_0)' \right)^{-1},$$

which follows from the ULLN, the continuity of $\mathbb{E} Z_t(\gamma) Z_t(\gamma)'$ and the consistency of $\vec{\gamma}$;

$$\text{(ii)} \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t(\vec{\gamma}) \varepsilon_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t(\gamma_0) \varepsilon_t + o_P(1),$$

which follows from Lemma C.2; (iii)$\frac{1}{T} \sum_{t=1}^{T} Z_t(\vec{\gamma}) x_t' \delta_0 (1_t(\vec{\gamma}) - 1_t) = o_P\left(T^{-1/2}\right)$, which also follows from Lemma C.2 and $\mathbb{E} |Z_t(\gamma) x_t' \delta_0 (1_t(\gamma) - 1_t)| \leq O(T^{-\varphi} |\gamma - \gamma_0|_2)$ as shown in claim 2. That is, we have shown that $\widehat{\alpha}(\vec{\gamma}) - \alpha_0 = \widehat{\alpha}(\gamma_0) - \alpha_0 + o_P\left(T^{-1/2}\right)$. ∎

Proof of claim 5. It can be proved using arguments identical to those used in Section C.1.3. ∎

D Proof of Selection Consistency in Section 5

Proof of Theorem 5.1. 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 2.1,

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

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 4.1 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.1 Selecting Relevant Factors via Iterative Estimation

In this subsection, we provide an detailed explanation of the iterative algorithm for selecting relevant factors in Section 5.

`[Iterative Estimation with Factor Selection]`

1. (Grid Construction) This step is the same as before.
2. (Initial Joint Estimation) This step is the same as before.
3. Iterate the following steps (a)-(c), beginning with $k = 1$ and terminating at a prespecified number $\bar{K}$.

   (a) For the given $\widehat{\alpha}^{k-1}$, obtain an estimate $\widehat{\gamma}^k$ via the mixed integer linear optimization algorithm

   $$\min_{\gamma\in\widetilde{\Gamma},\boldsymbol{d},\boldsymbol{e}} \frac{1}{T}\sum_{t=1}^{T} \left\{ (x_t'\widehat{\delta}^{k-1})^2 - 2(y_t - x_t'\widehat{\beta}^{k-1})x_t'\widehat{\delta}^{k-1} \right\} d_t + \lambda \sum_{m=1}^{p} e_m$$

   subject to (3.9) and (5.3).

   (b) For the given $\widehat{\gamma}^k$, obtain

   $$\widehat{\alpha}^k = \left[\frac{1}{T}\sum_{t=1}^{T} Z_t\left(\widehat{\gamma}^k\right) Z_t\left(\widehat{\gamma}^k\right)' \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} Z_t\left(\widehat{\gamma}^k\right) y_t$$

(c) Let $k = k + 1$.

(d) Finally, re-estimate the model with only selected factors.

In steps 1 and 2, it is necessary to use a grid for $\widetilde{\Gamma}$ without factor selection; on the other hand, in step 3(a), factor selection is implemented via the ℓ_0-norm penalized estimation given the initial estimator of α_0. The following theorem establishes the factor selection consistency. Its proof is given in Section D.

Theorem D.1. *Let Assumptions 1, 2, 3, and 4 hold. Suppose $\lambda T \to \infty$. Let $\widetilde{\gamma}$ denote the estimator of γ_0 using the iterative procedure described above for any $\bar{K} \geq 1$. Then,*

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

Proof of Theorem D.1. For $\alpha = (\beta, \delta)$, let

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

We prove the theorem by proving the following claims.

Claim 1. $\widetilde{\gamma}^0 \xrightarrow{p} \gamma_0$ for the approximate estimate $\widetilde{\gamma}^0 = \arg\min_{\gamma \in \Gamma_T} \min_\alpha \mathbb{S}_T\left(\alpha, \gamma\right)$.

Claim 2. For a given γ, let

$$\widehat{\alpha}\left(\gamma\right) = \arg\min_{\alpha} \mathbb{S}_T\left(\alpha, \gamma\right).$$

Then, for any $\vec{\gamma} \xrightarrow{p} \gamma_0$,

$$T^{\varphi}\left(\widehat{\alpha}\left(\vec{\gamma}\right) - \alpha_0\right) = o_P\left(1\right).$$

Claim 3. For a given α, let

$$\widetilde{\gamma}\left(\alpha\right) = \arg\min_{\gamma \in \Gamma} \mathbb{S}_T\left(\alpha, \gamma\right) + \lambda|\gamma|_0$$

Then, for any $\vec{\alpha} = \alpha_0 + o_P\left(T^{-\varphi}\right)$,

$$\widetilde{\gamma}\left(\vec{\alpha}\right) - \gamma_0 = O_P\left(T^{-1+2\varphi}\right),$$

and with probability approaching one,

$$S(\widetilde{\gamma}\left(\vec{\alpha}\right)) = S_0.$$

Claim 4. For $\vec{\gamma} = \gamma_0 + O_P\left(T^{-1+2\varphi}\right)$, and $S(\vec{\gamma}) = S_0$ with probability approaching one,

$$\widehat{\alpha}\left(\vec{\gamma}\right) = \alpha_0 + O_P\left(\frac{1}{\sqrt{T}}\right).$$

Then, for our iterative estimates, we can easily note that $\widehat{\alpha}^0 = \widehat{\alpha}\left(\widetilde{\gamma}^0\right)$ fulfils the conditions for claim 2 and $\widetilde{\gamma}^1$ does for claim 3 as $\widetilde{\gamma}^1 = \widetilde{\gamma}\left(\widehat{\alpha}^0\right)$, while $\widehat{\alpha}^1$ fits to claim 4 as $\widehat{\alpha}^1 = \widehat{\alpha}\left(\widetilde{\gamma}^1\right)$. ∎

Proofs of Claims 1 and 2. The proofs of Claims 1 and 2 are the same as those given in Section C.2. ∎

Proof of Claim 3. Given $\alpha = \alpha_0 + o_P(T^{-\varphi})$, we divide the proof in the following steps.

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

Step 2. Show that for $B_C(\gamma_0) = \{|\gamma - \gamma_0|_2 \leq CT^{-(1-2\varphi)}\}$ for any $C > 0$,

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

Step 3. Show that for $\widetilde{\gamma}_1(\alpha) = \arg\min_{S(\gamma)=S_0} \mathbb{S}_T(\alpha,\gamma)$ and $\widetilde{\gamma}_2(\alpha) = \arg\min_{\gamma} \mathbb{S}_T(\alpha,\gamma)$,

$$|\widetilde{\gamma}_j(\alpha) - \gamma_0|_2 = O_P(T^{-(1-2\varphi)}), \quad j = 1, 2.$$

Step 4. Show that $\min_{\gamma:S(\gamma)=S_0} \mathbb{S}_T(\alpha,\gamma) \leq \min_{\gamma} \mathbb{S}_T(\alpha,\gamma) + O_P(T^{-1})$.

Step 5. Show that for $\Gamma_b := \{\gamma : S_0 \subset S(\gamma), S_0 \neq S(\gamma)\}$,

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

with probability approaching one, where

$$\widetilde{\mathbb{S}}_T(\alpha,\gamma) = \mathbb{S}_T\left(\alpha,\gamma\right) + \lambda|\gamma|_0.$$

Now suppose $S_0 \neq S(\widetilde{\gamma}(\alpha))$. Then by step 1, $\widetilde{\gamma}(\alpha) \in \Gamma_b$, then by step 5,

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

which contradicts with the definition of $\widetilde{\gamma}(\alpha) := \arg\min_{\gamma} \widetilde{\mathbb{S}}_T(\alpha,\gamma)$. Consequently, we must have $S_0 = S(\widetilde{\gamma}(\alpha))$. In addition, given $S_0 = S(\widetilde{\gamma}(\alpha))$, we have

$$\widetilde{\gamma}(\alpha) := \arg\min_{S(\gamma)=S_0} \widetilde{\mathbb{S}}_T(\alpha,\gamma) = \arg\min_{S(\gamma)=S_0} \mathbb{S}_T(\alpha,\gamma) = \widetilde{\gamma}_1(\alpha),$$

where $\widetilde{\gamma}_1(\alpha)$ is defined in step 3. Thus by step 3, $|\widetilde{\gamma}(\alpha)-\gamma_0|_2 = O_P(T^{-(1-2\varphi)})$. ∎

Proof of Step 1. Let

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

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

$$\widetilde{\mathbb{S}}_T(\alpha,\gamma) - \mathbb{S}(\alpha,\gamma) = o_P(1).$$

and due to $\alpha = \alpha_0 + o_P(1)$,

$$\mathbb{S}(\alpha,\widetilde{\gamma}(\alpha)) = \widetilde{\mathbb{S}}_T(\alpha,\widetilde{\gamma}(\alpha)) + o_P(1) \le \widetilde{\mathbb{S}}_T(\alpha,\gamma_0) + o_P(1) = \mathbb{S}(\alpha,\gamma_0) + o_P(1) = \sigma^2 + o_P(1).$$

On the other hand, for $\Gamma_a = \{\gamma : S_0 \nsubseteqq S(\gamma)\}$,

$$\min_{\gamma\in\Gamma_a} \mathbb{E}\left(\alpha' Z_t(\gamma) - \alpha_0' Z_t(\gamma_0)\right)^2 = o_P(1) + \min_{\gamma\in\Gamma_a} \mathbb{E}\left(\alpha_0' Z_t(\gamma) - \alpha_0' Z_t(\gamma_0)\right)^2 > 0.$$

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

Proof of Step 2. $\mathbb{S}_T(\alpha,\gamma_1) - \mathbb{S}_T(\alpha,\gamma_2) = A(\gamma_1,\gamma_2) + B(\gamma_1,\gamma_2) + C(\gamma_1,\gamma_2)$ where, due to $\alpha = \alpha_0 + o_P(T^{-\varphi})$, uniformly for $\gamma_1,\gamma_2 \in B_C(\gamma_0)$,

$$\begin{aligned}
A(\gamma_1,\gamma_2) &= \frac{2}{T}\sum_t x_t'\delta\varepsilon_t(1\{f_t'\gamma_2>0\} - 1\{f_t'\gamma_1>0\}) \\
&= O_P(T^{-1}) + O_P(T^{-2\varphi})[|\gamma_1-\gamma_0| + |\gamma_2-\gamma_0|] = O_P(T^{-1}); \\
B(\gamma_1,\gamma_2) &= \frac{1}{T}\sum_t x_t'\delta(1\{f_t'\gamma_2>0\} - 1\{f_t'\gamma_1>0\})[Z_t(\gamma_0) - Z_t(\gamma_1) + Z_t(\gamma_0) - Z_t(\gamma_2)]'\alpha_0 \\
&\le O_P(T^{-2\varphi})\frac{1}{T}\sum_t |x_t|_2^2|1\{f_t'\gamma_1>0\} - 1\{f_t'\gamma_2>0\}| \\
&= O_P(T^{-2\varphi})(|\gamma_1-\gamma_0| + |\gamma_2-\gamma_0|) + O_P(T^{-(1+\varphi)}) = O_P(T^{-1}); \\
C(\gamma_1,\gamma_2) &= \frac{1}{T}\sum_t x_t'\delta(1\{f_t'\gamma_2>0\} - 1\{f_t'\gamma_1>0\})[Z_t(\gamma_1) + Z_t(\gamma_2)]'(\alpha_0-\alpha) \\
&\le o_P(T^{-2\varphi})\frac{1}{T}\sum_t |x_t|_2^2|1\{f_t'\gamma_1>0\} - 1\{f_t'\gamma_2>0\}| = O_P(T^{-1}).
\end{aligned}$$

∎

Proof of Step 3. By definition,

$$\mathbb{S}_T(\alpha,\widetilde{\gamma}_j(\alpha)) \le \mathbb{S}_T(\alpha,\gamma_0), \quad j=1,2.$$

Therefore, the same proof of claim 3 of the iterative estimation method carries over, which yields $|\widetilde{\gamma}_j(\alpha) - \gamma_0|_2 = O_P(T^{-(1-2\varphi)}), \quad j = 1, 2.$ ∎

Proof of Step 4. This step follows immediately from steps 2 and 3. ∎

Proof of Step 5. Given step 4, the proof then follows from a very similar argument of Step 3 in the proof of Theorem 5.1. So we omit the details. ∎

Proof of Claim 4. Given that $\vec{\gamma} = \gamma_0 + O_P\left(T^{-1+2\varphi}\right)$ and $S(\vec{\gamma}) = S_0$, the proof is the same as that of Claim 4 in Appendix C.2 for the iterative estimation. So we omit the details. ∎

E Proof of Asymptotics in Section 6: Estimated f (Joint Approach)

Similar to the case of known factors, the estimators of (α, γ) are defined using two approaches: one is the joint approach based on the MIQP and the other is the iterative approach based on the MILP. We split the proofs into two parts: the case of the joint approach and that of the iterative approach. We give the proofs for the joint approach in this section and those for the iterative approach in the next section.

E.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 E.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 $\widetilde{\mathbb{S}}_T(\alpha, \gamma)$ 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 E.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 E.3. Then as $\widetilde{\mathbb{S}}_T\left(\widehat{\alpha},\widehat{\gamma}\right) - \widetilde{\mathbb{S}}_T\left(\alpha_0,\gamma_0\right) \leq 0$, the decomposition yields: for $\widehat{\phi} = H_T\widehat{\gamma}$,

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

where $\mathbb{R}_T(\alpha,\phi)$ is lower bounded by $C|\alpha-\alpha_0|_2^2$ uniformly. Then, Lemmas E.3 and E.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 E.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 E.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 E.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 E.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 E.7.2. Due to the random rotation matrix H_T incurred by the factor estimation, we prove an extended continuous mapping theorem in Lemma H.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).

E.2 Discussion on Assumption 9

We discuss the reasons why Assumption 9 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 $\Psi = \mathbb{P}(0 < \breve{g}_t'\phi_0 < |\breve{g}_t'(\phi-\phi_0)|)$. 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 9 (i)-(iv).

E.3 Consistency

E.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 E.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 E.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

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

where

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

$$\begin{aligned}
A_{t,3} &= V_T^{-1}\frac{1}{T}\sum_{s=1}^{T}(\widetilde{f}_{1s}-\widetilde{H}_T'g_{1s})\frac{1}{N}(e_s'e_t-\mathbb{E}e_s'e_t),\\
A_{t,4} &= V_T^{-1}\widetilde{H}_T'\frac{1}{TN}\sum_{s=1}^{T}g_{1s}(e_s'e_t-\mathbb{E}e_s'e_t),\\
A_{t,5} &= V_T^{-1}\frac{1}{T}\sum_{s=1}^{T}(\widetilde{f}_{1s}-\widetilde{H}_T'g_{1s})g_{1t}'\frac{1}{N}\sum_{i=1}^{N}\lambda_i e_{is},\\
A_{t,6} &= V_T^{-1}\widetilde{H}_T'\frac{1}{T}\sum_{s=1}^{T}g_{1s}g_{1t}'\frac{1}{N}\sum_{i=1}^{N}\lambda_i e_{is}.
\end{aligned}$$

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

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

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

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

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 $|V_T-V|_2<\lambda_{\min}(V)/2$,

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

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

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

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

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

$$
\begin{aligned}
T^6\mathbb{P}(b_3 > \lambda_{\min}(V)/9) &= T^6\mathbb{P}(|\frac{1}{T}\sum_t e_t e_t' - \mathbb{E}e_t e_t'|_2 > N\lambda_{\min}(V)/9) \\
&\leq CT^{-3}\max_{ij}\mathbb{E}|\frac{1}{\sqrt{T}}\sum_t (e_{it}e_{jt} - \mathbb{E}e_{it}e_{jt})|^r \\
&\leq CT^{-3}\max_{ij}\int_0^\infty \mathbb{P}(|\frac{1}{\sqrt{T}}\sum_t (e_{it}e_{jt} - \mathbb{E}e_{it}e_{jt})| > x^{-r})dx \\
&\leq CT^{-3}\int_0^\infty \exp(-Cx^{-\zeta})dx = O(T^{-3}).
\end{aligned}
$$

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 $C_H > C_\lambda^2 C_v(2M_f)^{1/2}K$ (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 $c = 2c_1^{-1}$,

$$
\begin{aligned}
T^6\mathbb{P}(a_1 > CN^{-1}\log^c T) &\leq T^6\mathbb{P}(C\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T |\frac{1}{\sqrt{N}}(e_s'e_t - \mathbb{E}e_s'e_t)|^2 > C\log^c T) + o(1) \\
&\leq T^6\mathbb{P}(C\max_{st}|\frac{1}{\sqrt{N}}(e_s'e_t - \mathbb{E}e_s'e_t)|^2 > C\log^c T) + o(1) \\
&\leq T^8\max_{st}\mathbb{P}(|\frac{1}{\sqrt{N}}(e_s'e_t - \mathbb{E}e_s'e_t)| > C\log^{c/2} T) \\
&\leq C\exp(11\log T - C_1 C^{c_1}\log T) = o(1), \qquad \text{(E.5)}
\end{aligned}
$$

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

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

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 (E.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 (E.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,

$$\begin{aligned}
T^6 \max_t \mathbb{P}(|\frac{1}{\sqrt{TN}}\sum_{s=1}^T g_{1s}(e_s'e_t - \mathbb{E}e_s'e_t)|_2 > x^{1/2}) &\leq C\exp(10\log T - C_1 x^{c_1/2}) = o(1), \\
T^6\mathbb{P}(|\frac{1}{TN}\sum_{s=1}^T g_{1s}u_s'\Lambda|_2^2 > (NT)^{-1}y) &\leq C\exp(10\log T - C_1 y^{c_1/2}) = o(1), \\
T^6\mathbb{P}(\max_s |g_{1s}|_2^2 > z) &\leq \exp(6\log T - C_1 z^{c_1/2}) = o(1).
\end{aligned} \tag{E.7}$$

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

$$T^6\mathbb{P}(d_1 > (NT)^{-1}x) \leq T^6\mathbb{P}(\frac{1}{T}\sum_{t=1}^T |\frac{1}{TN}\sum_{s=1}^T g_{1s}(e_s'e_t - \mathbb{E}e_s'e_t)|_2^2 > (NT)^{-1}x)$$

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

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

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

Step 6: finishing the proof

We now work with (E.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 $Q < C m_{NT}^2$ with probability at least $1 - o(T^{-9})$. In addition,

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

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 $C > 0$,

$$
\begin{aligned}
T^6\mathbb{P}(|A_{t,1}|_2 > CT^{-1}(\log T)^{1/c_1}) & \leq T^6\mathbb{P}(\max_s |g_{1s}|_2 \sum_{s=1}^T |\frac{1}{N}\mathbb{E}e_s'e_t| > C(\log T)^{1/c_1}) + o(1) \\
& \leq T^6\mathbb{P}(\max_s |g_{1s}|_2 > C(\log T)^{1/c_1}) + o(1) = o(1), \\
T^6\mathbb{P}(|A_{t,2}|_2 > m_{NT}T^{-1/2}C) & \leq T^6\mathbb{P}(|\frac{1}{T}\sum_{s=1}^T(\widetilde{f}_{1s} - \widetilde{H}_T' g_{1s})\frac{1}{N}\mathbb{E}e_s'e_t|_2 > m_{NT}T^{-1/2}C) \\
& \leq T^6\mathbb{P}(Q\frac{1}{T}\sum_s |\frac{1}{N}\mathbb{E}e_s'e_t|^2 > m_{NT}^2 T^{-1}C^2) \\
& \leq T^6\mathbb{P}(\max_{st} |\frac{1}{N}\mathbb{E}e_s'e_t| \sum_s |\frac{1}{N}\mathbb{E}e_s'e_t| > C^2) + o(1) = o(1),
\end{aligned}
$$

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

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

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

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

∎

E.3.2 Defining notation

In the sequel, we show that $(\widehat{\alpha}, \widehat{\gamma})$ defined in Section 6.2 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 $1_t = 1\left\{f_t'\gamma_0 > 0\right\}$ 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:

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

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

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

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

Then we have

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

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.

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

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

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

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 E.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

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

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

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

where the first inequality is by the fact that $1\{A\}1\{B\}\leq 1\{A\}$ 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 $\mathbb{E}\left(|\varepsilon_t x_t|_2\,|g_t,h_t\right)\leq C$ a.s. in Assumption 5, and Proposition E.1.

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

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

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

(iii) The same proof as in part (i) leads to $\left|\widetilde{\mathbb{C}}_4(\delta,\gamma)-\widehat{\mathbb{C}}_4(\delta,\gamma)\right|\leq|\alpha-\alpha_0|_2O_P(n_{NT}+T^{-6})T^{-\varphi}$. 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),

$$|\widetilde{R}_3(\alpha,\gamma)-\widehat{R}_3(\alpha,\gamma)|\leq|\alpha-\alpha_0|_2T^{-\varphi}O_P(\Delta_f+T^{-6})$$

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

$$(\Delta_f+T^{-6})[T^{-2\varphi}+|\alpha-\alpha_0|_2^2+|\alpha-\alpha_0|_2T^{-\varphi}]\leq2(\Delta_f+T^{-6})[T^{-2\varphi}+|\alpha-\alpha_0|_2^2].$$

∎

E.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 E.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 H.2, uniformly in ϕ

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

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

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

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

(iv) Uniformly in ϕ,

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

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

Proposition E.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

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

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

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

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

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

For the first term in (E.10), recall $\breve{g}_t = g_t + h_t N^{-1/2}$ and note that by Lemma E.1, Lemma E.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 5 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 E.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 (E.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 E.1 and FCLT. Applying the same reasoning for the third term in (E.10) and recalling that $P(\gamma_0)X_0 = X_0$,

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

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,

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

by uniform law of large numbers. Also,

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

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

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

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

Since $\widehat{\phi}:=H_T\widehat{\gamma}$, Lemma E.1, E.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

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

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). ∎

E.4 Rate of convergence for $\widehat{\phi}$ (Proof of Theorem 6.1)

Here, we prove Theorem 6.1. Let

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

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

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

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

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

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 E.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 H.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 H.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 $\eta > 0$

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

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

$$\mathbb{C}_1(\delta_0, \phi) \leq T^{-2\varphi} O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right) + \eta T^{-2\varphi}\sqrt{N}\left|\phi - \phi_0\right|^2, \tag{E.11}$$

due to Lemma H.3. ∎

Lemma E.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 E.6.

Lemma E.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 $\mathbb{E}\mathbb{R}_2(\phi) = \mathbb{E}(x_t'\delta_0)^2\left|1\left\{\breve{g}_t'\phi > 0\right\} - 1\left\{\breve{g}_t'\phi_0 > 0\right\}\right|$. The proof is similar to Step 1 of Proof of Lemma C.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,

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

Thus,

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

for some $c_2 > 0$ because

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

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

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

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 $a_t := |f_t|_2 1\{M - \frac{|h_t|_2}{\sqrt{N}} < |f_t|_2 < M + \frac{|h_t|_2}{\sqrt{N}}\}$.

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

∎

Proposition E.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 E.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 $\mathbb{R}_3(\alpha, \phi) = \widehat{R}_3\left(\alpha, H_T^{-1}\phi\right)$, $\mathbb{R}_2(\phi) = \widehat{R}_2\left(H_T^{-1}\phi\right)$, $\mathbb{C}_i(\delta, \phi) = \widehat{\mathbb{C}}_i\left(\delta, H_T^{-1}\phi\right)$, $i = 1, 3$. In addition, since $\varphi < 1/2$, by Lemma E.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 E.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

$$\begin{aligned}&(C_0 - o_P(1))|\widehat{\alpha} - \alpha_0|_2^2 + \mathbb{G}_1(\widehat{\phi}) \leq \mathbb{G}_2(\widehat{\phi}) + \mathbb{C}_1(\delta_0, \widehat{\phi}) + O_P(\Delta_f + T^{-6})T^{-\varphi} \\ &+O_P(\Delta_f + T^{-1/2} + T^{-\varphi}N^{-1/2})|\widehat{\alpha} - \alpha_0|_2 + C_1 T^{-\varphi}\left|\widehat{\phi} - \phi_0\right|_2 |\widehat{\alpha} - \alpha_0|_2. \qquad \text{(E.12)}\end{aligned}$$

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, $|\widehat{\phi} - \phi_0|_2 \leq (C_2 C_3)/(8C_1^2)$. 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 E.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 E.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, (E.12) implies,

$$\begin{aligned} & C_0|\widehat{\alpha} - \alpha_0|_2^2/4 + C_4 T^{-2\varphi}|\widehat{\phi} - \phi_0|_2/2 \\ \leq \quad & O_P(T^{-1} + \frac{C}{\sqrt{N}T^{2\varphi}}) + O_P(\Delta_f + T^{-1/2} + T^{-\varphi}N^{-1/2})|\widehat{\alpha} - \alpha_0|_2 + O_P(\Delta_f + T^{-6})T^{-\varphi} \end{aligned} \tag{E.13}$$

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 E.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 E.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 E.6. *Let $u_t = g_t'\phi_0$ and Assumption 9 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,*

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

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 $A_{1t}(\phi)$ and note that (cf. notation $u_t = g_t'\phi_0$)

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

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

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

where

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

due to Assumption 9 (vi) for the first inequality. On the other hand, note that the normalization condition requires the first element of $\gamma-\gamma_0 = 0$, so $g_t'(\phi - \phi_0) = f_t'(\gamma-\gamma_0) = f_{2t}'(\gamma-\gamma_0)_2$. 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,

$$\begin{aligned}
&\mathbb{E}\left(x_t'\delta_0\right)^2 A_1 - A_{11} \\
&= \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{-h_t'\phi_0 < a \leq -\sqrt{N}g_t'\left(\phi-\phi_0\right) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 \leq 0\right\} p_{u_t|\bigstar}(\frac{a}{\sqrt{N}})da \\
&+ \frac{1}{\sqrt{N}}\mathbb{E}\left(x_t'\delta_0\right)^2 \int 1\left\{0 < a \leq -\sqrt{N}g_t'\left(\phi-\phi_0\right) - h_t'\phi_0\right\} 1\left\{h_t'\phi_0 > 0\right\} p_{u_t|\bigstar}(\frac{a}{\sqrt{N}})da \\
&= -\mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_t|\bigstar}(0) g_t'\left(\phi-\phi_0\right) 1\{g_t'\left(\phi-\phi_0\right) \leq 0\} 1\left\{h_t'\phi_0 \leq 0\right\} \\
&- \mathbb{E}\left(x_t'\delta_0\right)^2 p_{u_t|\bigstar}(0)\left(g_t'\left(\phi-\phi_0\right) + \frac{h_t'\phi_0}{\sqrt{N}}\right) 1\left\{g_t'\left(\phi-\phi_0\right) + \frac{h_t'\phi_0}{\sqrt{N}} < 0\right\} 1\left\{h_t'\phi_0 > 0\right\} \\
&+ B_1, \qquad \text{(E.14)}
\end{aligned}$$

?? where

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

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

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

due to Assumption 9 (vi).

Calculating A_2

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

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

So

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

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:

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

So

$$\mathbb{E}(x_t'\delta_0)^2 A_3$$

$$= \mathbb{E}(x_t'\delta_0)^2 1\left\{-h_t'\phi_0 < \sqrt{N}g_t'\phi_0 \le -\sqrt{N}g_t'(\phi-\phi_0) - h_t'\phi_0\right\} 1\{E_2\}$$
$$+\mathbb{E}(x_t'\delta_0)^2 1\left\{-h_t'\phi_0 < \sqrt{N}g_t'\phi_0 \le 0\right\} 1\{\sqrt{N}g_t'(\phi-\phi_0) + h_t'\phi_0 < 0\} + \mathbb{E}(x_t'\delta_0)^2 A_{31}.$$

Note that $\sqrt{N}\breve{g}_t = \sqrt{N}g_t + h_t$, so $|1\{E_3\} - 1\{E_2\}| \le 1\{E_5\} + 1\{E_6\}$. This gives, by Assumption 9 (vi) and letting $M_0 = 1$ to simplify the notation,

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

$$
\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2 A_3 - \mathbb{E}(x_t'\delta_0)^2 A_{31} \\
= \; & \mathbb{E}(x_t'\delta_0)^2 \int 1\left\{-h_t'\phi_0 < a \leq 0\right\} 1\{\sqrt{N} g_t'\left(\phi - \phi_0\right) + h_t'\phi_0 < 0\} p_{u_t|\bigstar}(\frac{a}{\sqrt{N}}) d\frac{a}{\sqrt{N}} \\
& + \mathbb{E}(x_t'\delta_0)^2 \int 1\left\{-h_t'\phi_0 < a \leq -\sqrt{N} g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\} 1\{g_t'\left(\phi - \phi_0\right) + \frac{h_t'\phi_0}{\sqrt{N}} > 0\} \\
& p_{u_t|\bigstar}(\frac{a}{\sqrt{N}}) d\frac{a}{\sqrt{N}} \\
= \; & \mathbb{E}(x_t'\delta_0)^2 \int 1\left\{-h_t'\phi_0 < a \leq -\sqrt{N} g_t'\left(\phi - \phi_0\right) - h_t'\phi_0\right\} 1\{g_t'\left(\phi - \phi_0\right) + \frac{h_t'\phi_0}{\sqrt{N}} > 0\} p_{u_t|\bigstar}(0) d\frac{a}{\sqrt{N}} \\
& + \mathbb{E}(x_t'\delta_0)^2 \int 1\left\{-h_t'\phi_0 < a \leq 0\right\} 1\{\sqrt{N} g_t'\left(\phi - \phi_0\right) + h_t'\phi_0 < 0\} p_{u_t|\bigstar}(0) d\frac{a}{\sqrt{N}} - B_3 \\
= \; & -\mathbb{E} p_{u_t|\bigstar}(0)(x_t'\delta_0)^2 g_t'\left(\phi - \phi_0\right) 1\{\frac{h_t'\phi_0}{\sqrt{N}} > g_t'\left(\phi - \phi_0\right) + \frac{h_t'\phi_0}{\sqrt{N}} > 0\} \\
& + \mathbb{E} p_{u_t|\bigstar}(0)(x_t'\delta_0)^2 \frac{h_t'\phi_0}{\sqrt{N}} 1\{\sqrt{N} g_t'\left(\phi - \phi_0\right) + h_t'\phi_0 < 0\} 1\{h_t'\phi_0 > 0\} - B_3
\end{aligned} \tag{E.15}
$$

where,

$$
\begin{aligned}
& |B_3| \leq \mathbb{E}(x_t'\delta_0)^2 \int 1\left\{-h_t'\phi_0 < a \leq 0\right\} 1\{\sqrt{N} g_t'\left(\phi - \phi_0\right) + h_t'\phi_0 < 0\} [p_{u_t|\bigstar}(\frac{a}{\sqrt{N}}) - p_{u_t|\bigstar}(0)] d\frac{a}{\sqrt{N}} \\
& + C|\mathbb{E}(x_t'\delta_0)^2 \frac{1}{N} \int 1\left\{-h_t'\phi_0 < a \leq [-\sqrt{N} g_t'\left(\phi - \phi_0\right) - h_t'\phi_0]\right\} 1\{g_t'\left(\phi - \phi_0\right) > -\frac{h_t'\phi_0}{\sqrt{N}}\} |a| da \\
\leq \; & \frac{C}{N} \mathbb{E}(x_t'\delta_0)^2 (|h_t'\phi_0| + |\sqrt{N} g_t'\left(\phi - \phi_0\right)|)^2 \leq \frac{C}{N T^{2\varphi}}.
\end{aligned}
$$

Calculating A_4

Write

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

The same proof as that of A_3 shows

$$
\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2 A_4 \\
= \; & \mathbb{E}(x_t'\delta_0)^2 (-h_t'\phi_0) 1\{h_t'\phi_0 < 0\} 1\{g_t'\left(\phi - \phi_0\right) + \frac{h_t'\phi_0}{\sqrt{N}} > 0\} p_{u_t|\bigstar(0)} \frac{1}{\sqrt{N}}
\end{aligned}
$$

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

Combining the above results, we reach,

$$\mathbb{E}(x_t'\delta_0)^2(A_1-A_3+A_2-A_4)=\sum_{d=1}^{8}\mathbb{E}[(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)a_d]+O(\frac{1}{T^{2\varphi}N^{0.5+1/(2p)}}) \tag{E.16}$$

where

$$\begin{aligned}
a_1 &= -\left(g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}<0\right\}1\left\{h_t'\phi_0>0\right\}\\
a_2 &= -g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)\le 0\}1\left\{h_t'\phi_0\le 0\right\}\\
a_3 &= g_t'(\phi-\phi_0)1\{\frac{h_t'\phi_0}{\sqrt{N}}>g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\}\\
a_4 &= -\frac{h_t'\phi_0}{\sqrt{N}}1\{\sqrt{N}g_t'(\phi-\phi_0)+h_t'\phi_0<0\}1\{h_t'\phi_0>0\}\\
a_5 &= g_t'(\phi-\phi_0)1\left\{g_t'(\phi-\phi_0)>0\right\}1\{h_t'\phi_0>0\}\\
a_6 &= \left(g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\right\}1\left\{h_t'\phi_0\le 0\right\}\\
a_7 &= \frac{h_t'\phi_0}{\sqrt{N}}1\{h_t'\phi_0<0\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}>0\}\\
a_8 &= -g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)>0\}1\{g_t'(\phi-\phi_0)+\frac{h_t'\phi_0}{\sqrt{N}}<0\}.
\end{aligned} \tag{E.17}$$

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

$$-\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)g_t'(\phi-\phi_0)1\{g_t'(\phi-\phi_0)\le 0\}1\left\{h_t'\phi_0\le 0\right\} \tag{E.18}$$
$$\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)g_t'(\phi-\phi_0)1\left\{g_t'(\phi-\phi_0)>0\right\}1\{h_t'\phi_0>0\}. \tag{E.19}$$

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 (E.18) and (E.19) 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 (E.18), we have by Assumption 8 and 9

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

Hence (E.18) 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, (E.19) 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

$$\begin{aligned}
& \mathbb{E}(x_t'\delta_0)^2(A_1 - A_3 + A_2 - A_4) \\
= & \mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)(a_1 + a_2' + a_3 + a_4 + a_5' + a_6 + a_7 + a_8). \\
= & -2\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)\left(g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}} < 0\right\}1\{h_t'\phi_0 > 0\} \\
& + 2\mathbb{E}(x_t'\delta_0)^2 p_{u_t|\bigstar}(0)\left(g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}}\right)1\left\{g_t'(\phi-\phi_0) + \frac{h_t'\phi_0}{\sqrt{N}} > 0\right\}1\{h_t'\phi_0 \leq 0\}.
\end{aligned} \tag{E.20}$$

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 9,

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

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 E.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 E.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

$$|\widehat{\phi}-\phi_0|_2 \leq O_P(\frac{1}{T^{1-2\varphi}}). \tag{E.21}$$

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 E.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 E.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 (E.12), 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 E.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 $T^{-1} = O(m_{NT})$ when $\sqrt{N} = O(T^{1-2\varphi})$.

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

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

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

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 (E.21), we reach

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

∎

E.5 Consistency of Regime Classification (Proof of Theorem 6.2)

Proof of Theorem 6.2. 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 H.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

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widetilde{f}_t'\gamma > 0\right\} - 1\left\{g_t'\phi_0 > 0\right\}\right| \leq \;& \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma > 0\right\} - 1\left\{\widetilde{f}_t'\gamma > 0\right\}\right| \\
& + \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma_0 > 0\right\} - 1\left\{\widehat{f}_t'\gamma > 0\right\}\right| \\
& + \frac{1}{T}\sum_{t=1}^{T}\left|1\left\{\widehat{f}_t'\gamma_0 > 0\right\} - 1\left\{g_t'\phi_0 > 0\right\}\right|.
\end{aligned} \tag{E.22}$$

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 (E.22), note that

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

and that

$$\begin{aligned} & \sup_{|\gamma|_2\leq C}\frac{1}{T}\sum_{t=1}^{T}1\{\widehat{f}_t'\gamma<0<\widetilde{f}_t'\gamma\} \qquad \text{(E.23)} \\ =\ & \sup_{|\gamma|_2\leq C}\frac{1}{T}\sum_{t=1}^{T}1\{-|\widehat{f}_t-\widetilde{f}_t|_2C<\widehat{f}_t'\gamma<0\} \\ \leq\ & \sup_{|\gamma|_2\leq C}\frac{1}{T}\sum_{t=1}^{T}1\left\{\left|\widehat{f}_t'\gamma\right|<C\Delta_f\right\}+\frac{1}{T}\sum_{t=1}^{T}1\{|\widehat{f}_t-\widetilde{f}_t|_2\geq\Delta_f\} \\ \leq\ & \frac{1}{T}\sum_{t=1}^{T}1\left\{\inf_{|\gamma|_2\leq C}\left|\widehat{f}_t'\gamma\right|<C\Delta_f\right\}+O_P(1)\,\mathbb{P}\{|\widehat{f}_t-\widetilde{f}_t|_2\geq\Delta_f\} \\ \leq\ & O_P(1)\,\mathbb{P}\left(\inf_{|\gamma|_2\leq C}|\widehat{f}_t'\gamma|<C\Delta_f\right)+O_P\left(T^{-6}\right) \\ \leq\ & O_P(\Delta_f+T^{-6}),\end{aligned}$$

where the first inequality is by the fact that $1\{A\}1\{B\}\leq 1\{A\}$ for any events A and B, and the remaining inequalities are by the law of iterated expectations, the rank condition in Assumption 5, and Proposition E.1. Recall in Proposition E.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

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

∎

Proof of Theorem 6.3 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.

E.6 Limiting distribution of $\widehat{\alpha}$ (Proof of Theorem 6.3: 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 E.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 E.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

$$r_{NT} := \left(NT^{1-2\varphi}\right)^{1/3} \wedge T^{1-2\varphi}. \tag{E.24}$$

Lemma E.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 E.7. (i) For each j,

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

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

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

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,

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

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,

$$|\widehat{f}_t|_2 \leq |H_T g_t|_2 + |H_T h_t|_2 N^{-1/2} < c(\log T)^c. \tag{E.25}$$

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 $\breve{g}_t=g_t+h_tN^{-1/2}$ and $\phi=H^{-1}\gamma$, and $\breve{g}_t$ is ρ-mixing. Since $\widehat{\phi}$ is consistent, by Lemma H.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 (E.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 (E.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 $\frac{1}{T}\sum_t[\widehat{Z}_t(\gamma_0) - Z_t(\gamma_0)]\varepsilon_t = o(T^{-1/2})$. ∎

E.7 Limiting distribution of $\widehat{\gamma}$ (Proof of Theorem 6.3: Part II)

Recall the defintion of r_{NT} in (E.24), 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:

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

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 E.8. *Let $\alpha = \alpha_0 + bT^{-1/2}$, and $\gamma = \gamma_0 + gr_{NT}^{-1}$. 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 E.1 and E.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 E.1 implies

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

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 E.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 E.1 and Lemma E.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 E.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 E.9 in the ensueing section. ∎

E.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

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

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 H.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 H.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 $c,\epsilon>0$

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

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 $v_t = \sqrt{r_{NT}}\varepsilon_t x_t' d_0 1\left\{-\breve{g}_t' \frac{\mathfrak{g}}{r_{NT}} < \breve{u}_t \leq 0\right\}$, 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, $T^{-2}\mathbb{E}\max_t v_t^4 \leq T^{-1}\mathbb{E}v_t^4$ by the stationarity and by $\max_t |a_t| \leq \sum_{t=1}^T |a_t|$ and $T^{-1}\mathbb{E}v_t^4 = T^{-1}r_{NT}^2\mathbb{E}\left(\varepsilon_t x_t' d_0\right)^4 1\left\{-\breve{g}_t' \frac{\mathfrak{g}}{r_{NT}} < \breve{u}_t \leq 0\right\} \leq CT^{-1}r_{NT} = o(1)$. Furthermore, $\frac{1}{T}\sum_{t=1}^T \left(v_t^2 - Ev_t^2\right) = o_P(1)$ due to Lemma H.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 E.9 below to conclude that

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

and that

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

Thus, we conclude that

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

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

Lemma E.9. *Assume Assumption 9. Then,*

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

as $N, T \to \infty$.

Proof of Lemma E.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,

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

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 9 (vii)) the convergence of $p_{\breve{u}_t}\left(0\right)$ 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 $\mathbb{E}\left(\eta_t^4\left|\breve{g}_t\right|_2^2\left|\breve{u}_t\right.\right)$.

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

$$\begin{aligned}
&\mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\left(\mathbb{E}\left(\eta_t^2|\frac{z}{r_{NT}},\breve{g}_t's,\breve{g}_t'w\right)-\mathbb{E}\left(\eta_t^2|0,\breve{g}_t's,\breve{g}_t'w\right)\right)\frac{z}{r_{NT}}p\left(0|\breve{g}_t's,\breve{g}_t'w\right)dz1\left\{\breve{g}_t's>\breve{g}_t'w\right\}\right] \\
\leq\ & C\mathbb{E}\left[\int_{-\breve{g}_t's}^{-\breve{g}_t'w}\frac{z}{r_{NT}}dzp\left(0|\breve{g}_t's,\breve{g}_t'w\right)1\left\{\breve{g}_t's>\breve{g}_t'w\right\}\right] \\
=\ & C'\mathbb{E}\left(\left(\breve{g}_t'w\right)^2-\left(\breve{g}_t's\right)^2\right)\frac{1}{2r_{NT}}=o\left(1\right).
\end{aligned}$$

∎

E.7.2 Bias

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

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

where

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

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

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

$$r_g=|\phi-\phi_0|_2^{-1}(\phi-\phi_0)=|Hg|_2^{-1}Hg.$$

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 E.6,

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

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 (E.17). 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:

$$\begin{aligned}
|a+b|-|b| &= \Xi(a,b), \quad \text{where} \\
\Xi(a,b) &:= -a1\{a\le 0\}1\{b\le 0\}-(a+b)1\{a+b<0\}1\{b>0\} \\
&\quad -b1\{a+b<0\}1\{b>0\}+a1\{a+b>0\}1\{a<0\} \\
&\quad +a1\{a>0\}1\{b>0\}+(a+b)1\{a+b>0\}1\{b\le 0\} \\
&\quad +b1\{b<0\}1\{a+b>0\}-a1\{a>0\}1\{a+b<0\}. \qquad \text{(E.26)}
\end{aligned}$$

Let $g_t'(\phi-\phi_0)=a$, $\frac{h_t'\phi_0}{\sqrt{N}}=b$, Note that (E.17) can be written exactly as the right hand side of the above equality, up to $\mathbb{E}_{|u_t=0}(x_t'\delta_0)^2p_{u_t}(0)$. Hence (E.17) and the above equality imply, for $\phi-\phi_0=r_{NT}^{-1}H_Tg$,

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

In the above, (1) is rewriting (E.17) 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 $\breve{\mathbb{C}}_{NT}(\mathfrak{g})=M_{NT}\mathbb{E}_{|u_t=0}(x_t'd_0)^2p_{u_t}(0)\mathbb{E}[\widetilde{F}_{g_t}(h_t'\phi_0)|x_t,g_t]$. 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

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

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 H.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

$$\breve{\mathbb{C}}_{NT}(\mathfrak{g}_T) \to \breve{A}(\mathfrak{g}). \tag{E.27}$$

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 (E.27), 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, (E.27) 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 $A\left(\omega, g\right) = A(\infty, g)$.

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 (E.20), 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

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

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

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

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 $\sigma_{h,x_t,g_t}^2 := \operatorname{plim}_{N\to\infty} \mathbb{E}[(h_t'\phi_0)^2|x_t, g_t, g_t'\phi_0 = 0]$. By

Assumption 8,

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

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

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

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

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

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

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

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,

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

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

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

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 H.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

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

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 (E.26) 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 (E.26).

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 $k \to 0$),

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

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: ($a = g_t' H g$)

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

$|(a+b) 1\{a+b<0\} 1\{b>0\} p_{\mathcal{Z}_t}(\zeta_k b)| < |(a+b) 1\{a+b<0\} 1\{b>0\}| C_0$, 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).

F Proof of Asymptotics in Section 6: Estimated f (Iterative Approach)

We now give the proofs for the iterative approach. We omit detailed discussions but sketch main differences from previous derivations in Section C.2 and E for the sake of space. Let

$$\widetilde{\mathbb{S}}_T(\gamma) = \min_{\alpha} \widetilde{\mathbb{S}}_T(\alpha, \gamma) = \min_{\alpha} \frac{1}{T} \sum_t (y_t - \widetilde{Z}_t(\gamma)'\alpha)^2.$$

Claim 1. $\widehat{\gamma}^0 \xrightarrow{p} \gamma_0$ for the approximate estimate $\widehat{\gamma}^0 = \operatorname{argmin}_{\gamma \in \Gamma_T} \widetilde{\mathbb{S}}_T(\gamma)$.

Claim 2. For a given γ, let

$$\widehat{\alpha}(\gamma) = \operatorname*{argmin}_{\alpha} \widetilde{\mathbb{S}}_T(\alpha, \gamma).$$

Then, for any $\gamma \xrightarrow{p} \gamma_0$,

$$T^{\varphi}(\widehat{\alpha}(\gamma) - \alpha_0) = o_P(1).$$

Claim 3. For a given α, let

$$\widehat{\gamma}(\alpha) = \operatorname*{argmin}_{\gamma \in \Gamma} \widetilde{\mathbb{S}}_T(\alpha, \gamma).$$

Then, for any $\vec{\alpha} = \alpha_0 + o_P(T^{-\varphi})$,

$$\widehat{\gamma}(\vec{\alpha}) - \gamma_0 = O_P\left(T^{-1+2\varphi} + N^{-1/2}\right),$$

Claim 4. For $\vec{\gamma} = \gamma_0 + O_P\left(T^{-1+2\varphi} + N^{-1/2}\right)$,

$$\widehat{\alpha}\left(\vec{\gamma}\right) = \widehat{\alpha}\left(\gamma_0\right) + o_P\left(\frac{1}{\sqrt{T}}\right),$$

and $\widehat{\alpha}\left(\gamma_0\right)$ is an oracle estimator:

$$\widehat{\alpha}\left(\gamma\right) - \alpha_0 \quad = \quad [\frac{1}{T}\sum_{t=1}^{T} Z_t\left(\gamma_0\right) Z_t\left(\gamma_0\right)']^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t\left(\gamma_0\right)\varepsilon_t + o_P(T^{-1/2}).$$

Claim 5. For $\alpha = \alpha_0 + O_P\left(T^{-1/2}\right)$,

$$\widehat{\gamma}(\alpha) - \gamma_0 = O_P(r_{NT}^{-1})$$

where

$$r_{NT}^{-1} = \max\left(\frac{1}{(NT^{1-2\varphi})^{1/3}}, \frac{1}{T^{1-2\varphi}}\right).$$

Claim 6. Derive the asymptotic independence of $r_{NT}\left(\widehat{\gamma}\left(\vec{\alpha}\right) - \gamma_0\right)$ and $\sqrt{T}\left(\widehat{\alpha}\left(\vec{\gamma}\right) - \alpha_0\right)$ and their marginal asymptotic distributions.

Then, for our iterative estimates, we can easily note that $\widehat{\alpha}^0 = \widehat{\alpha}\left(\widehat{\gamma}^0\right)$ fulfils the conditions for claim 2 and $\widehat{\gamma}^1$ does for claim 3 as $\widehat{\gamma}^1 = \widehat{\gamma}\left(\widehat{\alpha}^0\right)$, while $\widehat{\alpha}^1$ fits to claim 4 as $\widehat{\alpha}^1 = \widehat{\alpha}\left(\widehat{\gamma}^1\right)$. In addition, $\widehat{\gamma}^2$ fits to claim 5 as $\widehat{\gamma}^2 = \widehat{\gamma}(\widehat{\alpha}^1)$.

Proof of claim 1. It is sufficient if we show that $\widehat{\gamma}^0$ satisfies (E.9) in the proof of Proposition E.2, that is,

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

Repeating the argument using Lemma C.2 and the ULLN for the preceding derivation, we can observe that for any $c > 0$ there exists $T_0 < \infty$ such that for all $T > T_0$,

$$\begin{aligned}
& \widetilde{\mathbb{S}}_T\left(\widetilde{\gamma}\right) - \widetilde{\mathbb{S}}_T\left(\gamma_0\right) \\
\leq \quad & \max_{|\gamma-\gamma_0|\leq\psi_T}\left|\widetilde{\mathbb{S}}_T\left(\gamma\right) - \widetilde{\mathbb{S}}_T\left(\gamma_0\right)\right| \\
= \quad & \frac{1}{T}\max_{|\gamma-\gamma_0|\leq\psi_T}\left|e'\left(\widetilde{P}\left(\gamma_0\right) - \widetilde{P}\left(\gamma\right)\right)e + 2\delta_0'X_0\left(\widetilde{P}\left(\gamma_0\right) - \widetilde{P}\left(\gamma\right)\right)e + \delta_0'X_0'\left(\widetilde{P}\left(\gamma_0\right) - \widetilde{P}\left(\gamma\right)\right)X_0\delta_0\right| \\
\leq \quad & O_P\left(\frac{1}{T}\right) + O_P\left(\frac{T^{-\varphi}}{\sqrt{T}}\right) + o_P\left(T^{-2\varphi}\right) = o_P\left(T^{-2\varphi}\right),
\end{aligned}$$

where

$$\frac{1}{T}\delta_0'X_0'\left(\widetilde{P}\left(\gamma_0\right) - \widetilde{P}\left(\gamma\right)\right)X_0\delta_0 = O_P(T^{-2\varphi})[\frac{1}{T}X_0'(\widetilde{Z}(\gamma_0) - \widetilde{Z}(\gamma)) + \frac{1}{T}(\widetilde{Z}(\gamma)'\widetilde{Z}(\gamma) - \widetilde{Z}(\gamma_0)'\widetilde{Z}(\gamma_0))]$$

$$\leq O_P(T^{-2\varphi})\left[\Delta_f+T^{-6}+\sup_{|\gamma-\gamma_0|<\psi_T}\frac{1}{T}\sum_t |x_t|_2^2 1\{\widehat{f}_t\gamma>0\}-1\{\widehat{f}_t\gamma_0>0\}\right]=o_P\left(T^{-2\varphi}\right).$$

Proof of claim 2. Recall $1_t(\gamma)=1\{f_t'\gamma>0\}=1\{g_t'\phi>0\}$ for $\phi=H_T\gamma$; $1_t=1_t(\gamma_0)$.

$$\begin{aligned}
\widehat{\alpha}(\gamma)-\alpha_0 &= \left(\frac{1}{T}\sum_{t=1}^T \widetilde{Z}_t(\gamma)\widetilde{Z}_t(\gamma)'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \widetilde{Z}_t(\gamma)\varepsilon_t+\frac{1}{T}\sum_{t=1}^T\widetilde{Z}_t(\gamma)x_t'\delta_0\left(1\left(\widetilde{f}_t'\gamma>0\right)-1_t\right)\right)\\
&\leq O_P\left(\frac{1}{\sqrt{T}}+T^{-\varphi}(\Delta_f+T^{-6})\right)+O_P(1)\frac{1}{T}\sum_{t=1}^T\widehat{Z}_t(\gamma)x_t'\delta_0\left(1\left(\widehat{f}_t'\gamma>0\right)-1_t\right)\\
&\leq O_P\left(\frac{1}{\sqrt{T}}\right)+O_P(T^{-\varphi}\frac{1}{\sqrt{N}})+O_P(1)\frac{1}{T}\sum_{t=1}^T\widehat{Z}_t(\gamma)x_t'\delta_0\left(1\left(\widehat{f}_t'\gamma>0\right)-1\left(\widehat{f}_t'\gamma_0>0\right)\right)\\
&= O_P\left(\frac{1}{\sqrt{T}}\right)+O_P\left(T^{-2\varphi}|\gamma-\gamma_0|_2\right)+O_P(T^{-\varphi})\mathbb{E}\widehat{Z}_t(\gamma)x_t'd_0\left(1_t(\gamma)-1_t\right)\\
&= O_P\left(\frac{1}{\sqrt{T}}\right)+O_P\left(T^{-\varphi}|\gamma-\gamma_0|_2\right)=o_P(T^{-\varphi}). \qquad \text{(F.2)}
\end{aligned}$$

Proof of claim 3.

Note that for any γ,α,

$$\widetilde{\mathbb{S}}_T(\alpha,\gamma)=\widetilde{\mathbb{R}}_T(\alpha,\gamma)-\widetilde{\mathbb{G}}_T(\alpha,\gamma)+\text{terms independent of }\alpha,\gamma.$$

Recall the following quantities defined in Section E.3.2:

$$\begin{aligned}
\widetilde{\mathbb{R}}_T(\alpha,\gamma) &= \widetilde{R}_1(\alpha,\gamma)+\widetilde{R}_2(\gamma)+\widetilde{R}_3(\alpha,\gamma)\\
\widetilde{\mathbb{R}}_T(\alpha,\gamma_0) &= \widetilde{R}_1(\alpha,\gamma_0)\\
\widetilde{\mathbb{G}}_T(\alpha,\gamma) &= \widetilde{\mathbb{C}}_1(\alpha,\gamma)+\widetilde{\mathbb{C}}_2(\alpha)-\widetilde{\mathbb{C}}_3(\alpha,\gamma)-\widetilde{\mathbb{C}}_4(\alpha)\\
\widetilde{\mathbb{G}}_T(\alpha,\gamma_0) &= \widetilde{\mathbb{C}}_2(\alpha)-\widetilde{\mathbb{C}}_4(\alpha)
\end{aligned}$$

The rest of the proof is divided in the following steps.

claim 3: step i. consistency

First we show the consistency of $\widehat{\gamma}(\vec{\alpha})$ where $\vec{\alpha}=\alpha_0+o_P(T^{-\varphi})$. Note

$$\begin{aligned}
\widetilde{\mathbb{S}}_T(\alpha,\gamma)-\widetilde{\mathbb{S}}_T(\alpha,\gamma_0) &= \widetilde{R}_1(\alpha,\gamma)-\widetilde{R}_1(\alpha,\gamma_0)+\widetilde{R}_2(\gamma)+\widetilde{R}_3(\alpha,\gamma)-\widetilde{\mathbb{C}}_1(\alpha,\gamma)\\
&\quad+\widetilde{\mathbb{C}}_3(\alpha,\gamma). \qquad \text{(F.3)}
\end{aligned}$$

Now for any $\alpha=\alpha_0+o_P(T^{-\varphi})$, and $\gamma=\widehat{\gamma}(\alpha)$, $\widetilde{\mathbb{S}}_T(\alpha,\gamma)-\widetilde{\mathbb{S}}_T(\alpha,\gamma_0)\leq 0$.

$$T^{2\varphi}\sup_{\gamma}[\widetilde{R}_1(\alpha,\gamma)+|\widetilde{R}_3(\alpha,\gamma)|+|\widetilde{\mathbb{C}}_3(\alpha,\gamma)|]=o_P(1)$$

Also,

$$T^{2\varphi}\sup_{\gamma}|\widetilde{\mathbb{C}}_1(\alpha,\gamma)|\leq O_P(T^{-\varphi})T^{2\varphi}\sup_{\gamma}|\frac{4}{T}\sum_{t=1}^{T}\varepsilon_t x_t' 1\{\widetilde{f}_t'\gamma>0\}|_2=O_P(T^{\varphi})T^{-1/2}=o(1)$$

Also by Lemma E.1, $T^{2\varphi}\sup_{\gamma}|\widetilde{R}_2(\gamma)-\widehat{R}_2(\gamma)|=o_P(1)$ where

$$\widehat{R}_2(\gamma) \quad = \quad \frac{1}{T}\sum_{t=1}^{T}\left(x_t'\delta_0\right)^2\left|1\left\{\widehat{f}_t'\gamma>0\right\}-1\left\{\widehat{f}_t'\gamma_0>0\right\}\right|.$$

By lemma C.2, uniformly in γ, $T^{2\varphi}|\widehat{R}_2(\gamma)-\mathbb{E}\widehat{R}_2(\gamma)|\leq[O_P(T^{-(1-\varphi)})+\eta T^{-\varphi}|\gamma-\gamma_0|]$. Also, $T^{2\varphi}\mathbb{E}\widehat{R}_2(\gamma)=T^{2\varphi}\mathbb{E}\left(x_t'd_0\right)^2\left|1\left\{\widehat{f}_t'\gamma>0\right\}-1\left\{\widehat{f}_t'\gamma_0>0\right\}\right|\geq c|\gamma-\gamma_0|-o_P(1)$. We then reach

$$(c-\eta T^{-\varphi})|\gamma-\gamma_0|_2+o_P(1)\leq 0$$

leading to the consistency of γ.

claim 3: step ii. rate of convergence

We now study each term on the right of (F.3).

(i) $\widetilde{R}_1(\alpha,\gamma)-\widetilde{R}_1(\alpha,\gamma_0)$**.** By lemma E.1 and E.2, uniformly in γ, and $\phi=H_T\gamma$,

$$\begin{aligned}\widetilde{R}_1(\alpha,\gamma) &= \widehat{R}_1(\alpha,\gamma)+[|\alpha-\alpha_0|_2^2+T^{-2\varphi}]O_P(\Delta_f+T^{-6})\\ &= (\alpha-\alpha_0)'\frac{1}{T}\sum_t\breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)'(\alpha-\alpha_0)+[|\alpha-\alpha_0|_2^2+T^{-2\varphi}]O_P(\Delta_f+T^{-6})\end{aligned}$$

Now by Lemma H.2, recall $\breve{g}_t=g_t+h_tN^{-1/2}$.

$$\begin{aligned}&(\alpha-\alpha_0)'[\frac{1}{T}\sum_t\breve{\mathbb{Z}}_t(\phi)\breve{\mathbb{Z}}_t(\phi)'-\frac{1}{T}\sum_t\breve{\mathbb{Z}}_t(\phi_0)\breve{\mathbb{Z}}_t(\phi_0)'](\alpha-\alpha_0)\\ \leq\ & C|\alpha-\alpha_0|_2^2|\frac{1}{T}\sum_t|x_t|_2^2[1\{\breve{g}_t'\phi<0<\breve{g}_t'\phi_0\}]+C|\alpha-\alpha_0|_2^2|\frac{1}{T}\sum_t|x_t|_2^2[1\{\breve{g}_t'\phi_0<0<\breve{g}_t'\phi\}]\\ \leq\ & C|\alpha-\alpha_0|_2^2\mathbb{E}|x_t|_2^2 1\{\breve{g}_t'\phi<0<\breve{g}_t'\phi_0\}+C|\alpha-\alpha_0|_2^2\mathbb{E}|x_t|_2^2 1\{\breve{g}_t'\phi_0<0<\breve{g}_t'\phi\}\\ & +|\alpha-\alpha_0|_2^2[\eta T^{-\varphi}|\phi-\phi_0|_2+O_P(T^{\varphi}T^{-1})]\\ \leq\ & C|\phi-\phi_0|_2|\alpha-\alpha_0|_2^2+|\alpha-\alpha_0|_2^2O_P(T^{-1+\varphi})\end{aligned}$$

Hence

$$|\widetilde{R}_1(\alpha,\gamma)-\widetilde{R}_1(\alpha,\gamma_0)|=|\alpha-\alpha_0|_2^2O_P(\Delta_f+T^{-6}+T^{-1+\varphi})+C|\phi-\phi_0|_2|\alpha-\alpha_0|_2^2+T^{-2\varphi}O_P(\Delta_f+T^{-6}).$$

(ii) $\widetilde{R}_3(\alpha,\gamma)$**.** By lemma E.1 and E.2, uniformly in γ,

$$\widetilde{R}_3(\alpha,\gamma) \quad \leq \quad [|\alpha-\alpha_0|_2^2 + T^{-2\varphi}]O_P(\Delta_f + T^{-6}) + \left(O_P\left(T^{-1}\right) + CT^{-\varphi}\left|\phi-\phi_0\right|_2\right)\left|\alpha-\alpha_0\right|_2.$$

(iii) $\widetilde{\mathbb{C}}_1(\alpha,\gamma)$**.** By Lemma E.1 and E.2,

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

(iv) $\widetilde{R}_2(\gamma) + \widetilde{\mathbb{C}}_3(\alpha,\gamma)$**.** Recall

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

By lemma E.1 and E.2, uniformly in γ, and $\phi = H_T\gamma$,

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

(v) Putting together. $\widetilde{\mathbb{S}}_T\left(\alpha,\gamma\right) - \widetilde{\mathbb{S}}_T\left(\alpha,\gamma_0\right) \leq 0$ implies

$$0 \geq \widetilde{R}_1(\alpha,\gamma) - \widetilde{R}_1(\alpha,\gamma_0) + \widetilde{R}_2(\gamma) + \widetilde{R}_3(\alpha,\gamma) - \widetilde{\mathbb{C}}_1(\alpha,\gamma) + \widetilde{\mathbb{C}}_3(\alpha,\gamma).$$

Then due to $|\alpha-\alpha_0|_2 = o_P(T^{-\varphi})$,

$$\begin{aligned}&\mathbb{G}_1(\phi) + \mathbb{G}_2(\phi) - \mathbb{C}_1\left(\delta_0,\phi\right) \\ \leq\ &\left(T^{-\varphi}N^{-1/2} + T^{-1+\varphi}\right)\left|\alpha-\alpha_0\right|_2 + T^{-\varphi}O_P(\Delta_f + T^{-6}) \\ &+(C+\eta)T^{-\varphi}\left|\phi-\phi_0\right|_2\left|\alpha-\alpha_0\right|_2 + |\alpha-\alpha_0|_2^2 O_P(T^{-1+\varphi}) \\ \leq\ &o_P(T^{-2\varphi})N^{-1/2} + o_P(T^{-1}) + T^{-\varphi}O_P(\Delta_f + T^{-6}) + o_P(T^{-2\varphi})\left|\phi-\phi_0\right|_2 \qquad \text{(F.4)}\end{aligned}$$

By Lemmas E.4, E.5, $|\mathbb{G}_2(\phi)| + |\mathbb{C}_1\left(\delta_0,\phi\right)| \leq b_{NT}$, and $\mathbb{G}_1(\phi) \geq CT^{-2\varphi}|\phi-\phi_0|_2 - \frac{C}{\sqrt{N}T^{2\varphi}}$, where for an arbitrarily small $\eta > 0$, $b_{NT} = O_P(T^{-1}) + \eta T^{-2\varphi}\left|\phi-\phi_0\right|_2$. Then

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

Since $\eta > 0$ is arbitrarily small, we have

$$|\phi - \phi_0|_2 \le O_P(N^{-1/2} + T^{-(1-2\varphi)}).$$

Proof of claim 4.

Write $\mathcal{A}_T(\gamma) = \frac{1}{T}\sum_{t=1}^T \widetilde{Z}_t(\gamma)\widetilde{Z}_t(\gamma)'$. By (F.2)

$$\begin{aligned}
\widehat{\alpha}(\gamma) - \alpha_0 &= \mathcal{A}(\gamma)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \widetilde{Z}_t(\gamma)\varepsilon_t + \frac{1}{T}\sum_{t=1}^T \widetilde{Z}_t(\gamma) x_t'\delta_0\left(1\left(\widetilde{f}_t'\gamma > 0\right) - 1_t\right)\right) \\
&= \mathcal{A}(\gamma)^{-1}\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma)\varepsilon_t + \mathcal{A}(\gamma)^{-1}\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma) x_t'\delta_0\left(1\left(\widehat{f}_t'\gamma > 0\right) - 1_t\right) + O_P(\Delta_f + T^{-6}) \\
&= \mathcal{A}(\gamma)^{-1}\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma)\varepsilon_t + \mathcal{A}(\gamma)^{-1}\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma) x_t'\delta_0\left(1\left(\widehat{f}_t'\gamma > 0\right) - 1\left(\widehat{f}_t'\gamma_0 > 0\right)\right) \\
&\quad + O_P(\Delta_f + T^{-6} + T^{-\varphi}N^{-1/2})
\end{aligned} \tag{F.5}$$

By the proof of lemma E.7,

$$\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma)\varepsilon_t = \frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma_0)\varepsilon_t + o_P(T^{-1/2}) = \frac{1}{T}\sum_{t=1}^T Z_t(\gamma_0)\varepsilon_t + o_P(T^{-1/2}).$$

On the other hand, by lemma H.2, uniformly in γ, since $T = O(N)$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T \widehat{Z}_t(\gamma) x_t'\delta_0\left(1\left(\widehat{f}_t'\gamma > 0\right) - 1\left(\widehat{f}_t'\gamma_0 > 0\right)\right) \\
=\ & \mathbb{E}\widehat{Z}_t(\gamma) x_t'\delta_0\left(1\left(\widehat{f}_t'\gamma > 0\right) - 1\left(\widehat{f}_t'\gamma_0 > 0\right)\right) + \eta T^{-2\varphi}|\gamma - \gamma_0|_2 + O_P(T^{-1}) \\
\le\ & O(T^{-\varphi})|\gamma - \gamma_0|_2 + O_P(T^{-1}) \\
\le\ & O(N^{-1/2}T^{-\varphi} + T^{-1+\varphi}) + O_P(T^{-1}) = o_P(T^{-1/2}).
\end{aligned} \tag{F.6}$$

So

$$\begin{aligned}
\widehat{\alpha}(\gamma) - \alpha_0 &= \mathcal{A}(\gamma)^{-1}\frac{1}{T}\sum_{t=1}^T Z_t(\gamma_0)\varepsilon_t + o_P(T^{-1/2}) \\
&= [\frac{1}{T}\sum_{t=1}^T Z_t(\gamma_0) Z_t(\gamma_0)']^{-1}\frac{1}{T}\sum_{t=1}^T Z_t(\gamma_0)\varepsilon_t + o_P(T^{-1/2}).
\end{aligned} \tag{F.7}$$

This immediately implies $\widehat{\alpha}(\gamma) - \widehat{\alpha}(\gamma_0) = o_P(T^{-1/2})$.

Proof of claim 5.

In claim 3, we proved $\widehat{\gamma}(\alpha) - \gamma_0 = O_P(N^{-1/2} + T^{-(1-2\varphi)})$. Now suppose $\sqrt{N} = O(T^{1-2\varphi})$.

By lemmas E.4, E.5, E.6, for $\phi = H_T\widehat{\gamma}(\alpha)$,

$$|\mathbb{G}_2(\phi)| + |\mathbb{C}_1(\delta_0, \phi)| \leq a_{NT}, \quad \mathbb{G}_1(\phi) \geq CT^{-2\varphi}\sqrt{N}|\phi - \phi_0|_2^2 - O(\frac{1}{T^{2\varphi}N^{5/6}}),$$

where for an arbitrarily small $\eta > 0$, $a_{NT} = T^{-2\varphi}O_P\left(\frac{\sqrt{N}}{(NT^{1-2\varphi})^{2/3}}\right) + T^{-2\varphi}\eta|\phi - \phi_0|_2^2\sqrt{N}$. Then due to $\alpha = \alpha_0 + O_P\left(T^{-1/2}\right)$, (F.4) implies

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

Combining with the rates proved in claim 3, we obtain the desired result.

Proof of claim 6.

Let $l_{NT} = \sqrt{r_{NT}T^{1+2\varphi}}$ and $g = r_{NT}(\gamma - \gamma_0)$. We have

$$\begin{aligned} l_{NT}\left(\widetilde{\mathbb{S}}_T(\alpha,\gamma) - \widetilde{\mathbb{S}}_T(\alpha,\gamma_0)\right) &= l_{NT}[\widetilde{R}_1(\alpha,\gamma) - \widetilde{R}_1(\alpha,\gamma_0)] + l_{NT}\widetilde{R}_2(\gamma) + l_{NT}\widetilde{R}_3(\alpha,\gamma) \\ &\quad -l_{NT}\widetilde{\mathbb{C}}_1(\alpha,\gamma) + l_{NT}\widetilde{\mathbb{C}}_3(\alpha,\gamma) \end{aligned} \tag{F.8}$$

For some $c > 0$, $\Delta_f = \log^c T/T$, so by the proof of claim 3,

$$\begin{aligned} l_{NT}|\widetilde{R}_1(\alpha,\gamma) - \widetilde{R}_1(\alpha,\gamma_0)| &= l_{NT}|\alpha - \alpha_0|_2^2 O_P(T^{-1+\varphi} + |\phi - \phi_0|_2) + l_{NT}T^{-2\varphi}O_P(\Delta_f + T^{-6}) \\ &= O_P(\frac{1}{T^{1/2-\varphi}r_{NT}^{1/2}} + \frac{r_{NT}^{1/2}\log^c T}{T^{1/2+\varphi}}) = o_P(1). \end{aligned}$$

By the proof of Lemma E.8, $l_{NT}\mathbb{G}_2 = o_P(1)$, and

$$\begin{aligned} l_{NT}|\widetilde{R}_3| + l_{NT}|\widetilde{R}_2(\gamma) - \mathbb{R}_2(\phi)| &\leq o_P(1) \\ l_{NT}|\widetilde{\mathbb{C}}_1(\delta,\gamma) - \widehat{\mathbb{C}}_1(\delta_0,\gamma)| + l_{NT}\left|\widehat{\mathbb{C}}_3(\delta_0,\gamma) - \widetilde{\mathbb{C}}_3(\delta,\gamma)\right| &\leq o_P(1). \end{aligned}$$

Hence

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

By the continuous mapping theorem for the argmin function,

$$\begin{aligned} r_{NT}(\widehat{\gamma}(\alpha) - \gamma_0) &= \arg\min_g l_{NT}\left(\widetilde{\mathbb{S}}_T\left(\alpha,\gamma_0 + gr_{NT}^{-1}\right) - \widetilde{\mathbb{S}}_T(\alpha,\gamma_0)\right) \\ &= \arg\min_g l_{NT}\mathbb{E}[\widehat{R}_2(\gamma_0 + gr_{NT}^{-1}) + \widehat{\mathbb{C}}_3(\alpha_0,\gamma_0 + gr_{NT}^{-1})] \\ &\quad -l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0,\gamma_0 + gr_{NT}^{-1}\right) + o_P(1). \end{aligned}$$

Given this, it then follows from the proof of Theorem 6.3 that

$$r_{NT}\left(\widehat{\gamma}(\alpha)-\gamma_0\right) \xrightarrow{d} \underset{g\in\mathcal{G}}{\operatorname{argmin}}\, A\left(\omega,g\right)+2W\left(g\right).$$

Finally, the above result also holds when $\widehat{\gamma}(\alpha)$ is replaced with $\widehat{\gamma}(\alpha_0)$ by setting $\alpha=\alpha_0$. More specifically,

$$\begin{aligned} r_{NT}\left(\widehat{\gamma}(\alpha_0)-\gamma_0\right) = \arg\min_{g} l_{NT}\mathbb{E}[\widehat{R}_2(\gamma_0+gr_{NT}^{-1})+\widehat{\mathbb{C}}_3(\alpha_0,\gamma_0+gr_{NT}^{-1})] \\ - l_{NT}\widehat{\mathbb{C}}_1\left(\delta_0,\gamma_0+gr_{NT}^{-1}\right)+o_P(1). \end{aligned}$$

Taking the difference yields

$$r_{NT}[\widehat{\gamma}(\alpha)-\widehat{\gamma}(\alpha_0)] = o_P(1).$$

G Proof of Linearity Test in Section 7

Proof of Theorem 7.1. We begin with the known factor case. For each γ, our $Q_T\left(\gamma\right)$ corresponds to a modified version of the Wald statistic $T_n\left(\gamma\right)$ 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

$$\min_{\alpha:\delta=0}\mathbb{S}_T\left(\alpha,\gamma\right)-\min_{\alpha,\gamma}\mathbb{S}_T\left(\alpha,\gamma\right)=\widehat{\alpha}(\gamma)'R'[R(\sum_t Z_t(\gamma)Z_t(\gamma)')^{-1}R']^{-1}R\widehat{\alpha}(\gamma).$$

We then replace the term $\widehat{V}_n\left(\gamma\right)$ in Hansen (1996) with

$$\widehat{V}_n\left(\gamma\right)=\frac{1}{T}\sum_{t=1}^{T}x_tx_t'1\left\{f_t'\gamma>0\right\}\mathbb{S}_T\left(\widehat{\alpha},\widehat{\gamma}\right). \tag{G.1}$$

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 (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 $T^{-1/2}\sum_{t=1}^{T}x_t1\left\{f_t'\gamma>0\right\}\varepsilon_t$, which is guaranteed by our maximal inequality Lemma H.1. Finally, his Assumption 3 follows from the ULLN. Then, the theorem is proved with the replaced $\widehat{V}_n\left(\gamma\right)$ in (G.1).

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

$$\sup_{\gamma}\left|\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\left(1\left\{\widetilde{f}_t'\gamma>0\right\}-1\left\{f_t'\gamma>0\right\}\right)\right|=o_P(1), \tag{G.2}$$

$$\sup_{\gamma}\left|\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\left(1\left\{\widetilde{f}_t'\gamma>0\right\}-1\left\{f_t'\gamma>0\right\}\right)\varepsilon_t^2\right|=o_P(1), \tag{G.3}$$

$$\sup_{\gamma}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t\left(1\left\{\widetilde{f}_t'\gamma>0\right\}-1\left\{f_t'\gamma>0\right\}\right)\varepsilon_t\right|=o_P(1). \tag{G.4}$$

Recall that $\widehat{f}_t$ is defined as $\widehat{f}_t = H_T'(g_t + h_t/\sqrt{N})$. The last condition (G.4) follows directly if we show that

$$\sup_{\gamma}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t\left(1\left\{\widetilde{f}_t'\gamma>0\right\}-1\left\{\widehat{f}_t'\gamma>0\right\}\right)\varepsilon_t\right|=o_P(1) \tag{G.5}$$

and

$$\sup_{\gamma}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t\left(1\left\{\widehat{f}_t'\gamma>0\right\}-1\left\{f_t'\gamma>0\right\}\right)\varepsilon_t\right|=o_P(1). \tag{G.6}$$

By Lemma E.1, (G.5) follows. To show (G.6), note that in view of the maximal inequality in Lemma H.1 and Theorem 16.1 of Billingsley (1968), 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 (G.6). The other two conditions (G.2) and (G.3) can be shown similarly and thus omitted. ∎

H 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 10. *Assume that $\{z_t, q_t\}_{t=1}^T$ be a sequence of strictly stationary, ergodic, and ρ-mixing array with $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$, $\mathbb{E}\,|z_t|_2^4<\infty$, and, for all γ in a neighborhood of γ_0, $\mathbb{E}\left(|z_t|^4\,|q_t=\gamma\right)<C<\infty$ 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 H.1. *Let Assumption 10 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 $\epsilon \geq T^{-1}$,*

$$\mathbb{P}\left\{ \sup_{|\gamma-\vec{\gamma}|_2<\epsilon} \left| \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(z_t 1_t \left(\vec{\gamma}, \gamma \right) - \mathbb{E} z_t 1_t \left(\vec{\gamma}, \gamma \right) \right) \right| > K \right\} \leq \frac{C}{K^4} \epsilon^2.$$

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 H.2. *Let Assumption 10 hold. For any $\eta > 0$ and some $C > 0$,*

$$\sup_{T^{-1+2\varphi}<|\gamma-\gamma_0|_2<C} \left[\left| \frac{1}{T^{1+\varphi}} \sum_{t=1}^{T} \left(z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) - \mathbb{E} z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) \right) \right| \right.$$
$$\left. - \eta T^{-2\varphi} |\gamma - \gamma_0|_2 \right] \leq O_P\left(\frac{1}{T}\right).$$

Lemma H.3. *Let Assumption 10 hold. For any $\eta > 0$ and some $C > 0$,*

$$\sup_{T^{-1+2\varphi}<|\gamma-\gamma_0|_2<C} \left[\left| \frac{1}{\sqrt{N} T^{1-\varphi}} \sum_{t=1}^{T} \left(z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) - \mathbb{E} z_t \left(1_t(\gamma) - 1_t(\gamma_0) \right) \right) \right| - \eta |\gamma - \gamma_0|_2^2 \right]$$
$$\leq \quad O_P\left(\frac{1}{\left(NT^{1-2\varphi}\right)^{2/3}}\right).$$

We derive an extended continuous mapping theorem (CMT) in Lemma H.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 H.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

$$\mathbb{P}\left\{\left|f_n(z) - f_n(z')\right|_2 > C_\eta \left|z - z'\right|_2 \;\; \textit{for all } z, z'\right\} < \eta,$$

for all n. *Then,*

$$\mathbb{G}_n(f_n(z)) \Rightarrow \mathbb{G}(f(z))$$

over any compact set.

H.1 Proofs of Lemmas

Proof of Lemma H.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,

$$\begin{aligned} &\mathbb{E}\left|\frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(z_t 1_t(\gamma_1, \gamma_2) - \mathbb{E} z_t 1_t(\gamma_1, \gamma_2)\right)\right|^4 \\ &\le C\left(T^{-1}\mathbb{E}\left|z_t\right|^4 \; 1_t(\gamma_1, \gamma_2) + \left(\mathbb{E}\left|z_t\right|^2 \; 1_t(\gamma_1, \gamma_2)\right)^2\right). \end{aligned} \tag{H.1}$$

Consider $\gamma_1 = (\psi', c_1)'$ and $\gamma_2 = (\psi', c_2)'$, which are identical other than the last elements. Then,

$$1_t(\gamma_1, \gamma_2) = 1\left\{c_2 < q_{1t}'\psi \le c_1\right\}$$

and thus there is a universal constant C such that

$$\begin{aligned} \mathbb{E}\left|z_t\right|^k \; 1_t(\gamma_1, \gamma_2) &= \mathbb{E}\left[\mathbb{E}\left(\left|z_t\right|^k \middle| q_t\right) 1_t(\gamma_1, \gamma_2)\right] \\ &\le C\mathbb{E}1_t(\gamma_1, \gamma_2) \le C' \left|c_1 - c_2\right| \end{aligned}$$

for $k = 2, 4$, as the densities of $q_t'\gamma$ are bounded uniformly. Thus, for any c_1, c_2 such that $|c_1 - c_2| \ge T^{-1}$,

$$\sup_{\psi} \mathbb{E}\left|\frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(z_t 1_t(\gamma_1, \gamma_2) - \mathbb{E} z_t 1_t(\gamma_1, \gamma_2)\right)\right|^4 \le C\left|c_1 - c_2\right|^2. \tag{H.2}$$

Here, recall that ψ is the common element between γ_1 and γ_2.

Next, by Bickel and Wichura (1971), their equation (1), that

$$\sup_{\gamma} |J_T(\gamma)| \leq d \cdot M'' + |J_T(\widetilde{\gamma})|,$$

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[14] $B = (\gamma_1, \gamma_2] = (\gamma_{12,}\gamma_{22}] \times \cdots \times (c_1, c_2]$ with each side of length greater than equal to 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 (H.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

$$\mathbb{P}\{M'' > K\} \leq \frac{C}{K^4} \mu(T)^{2/d} \leq \frac{C}{K^4} \epsilon^2,$$

for some $C < \infty$.

Furthermore, the Markov inequality, the moment bound in (H.1), the boundedness of the

[14]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.

density of $q_t'\gamma$ imply that

$$\mathbb{P}\left\{\left|J_T\left(\widetilde{\gamma}\right)\right| > K\right\} \leq \frac{C}{K^4}\epsilon^2,$$

for some $C < \infty$. This completes the proof. ∎

Proof of Lemma H.2. Define $A_{T,j} = \{\theta : (j-1)T^{-1+2\varphi} \leq |\gamma - \gamma_0|_2 < jT^{-1+2\varphi}\}$ and

$$R_T^2 = T \sup_{T^{-1+2\varphi} < |\gamma - \gamma_0| \leq C} \left[\left|\mathbb{D}_T\left(\gamma\right)\right| - \eta T^{-2\varphi}\left|\gamma - \gamma_0\right|_2\right],$$

where $\mathbb{D}_T\left(\gamma\right) = \frac{1}{T^{1+\varphi}}\sum_{t=1}^{T}\left(z_t\left(1_t\left(\gamma\right) - 1_t\left(\gamma_0\right)\right) - \mathbb{E}z_t\left(1_t\left(\gamma\right) - 1_t\left(\gamma_0\right)\right)\right)$. Then, for any $m > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{R_T > m\right\} \\
&= \mathbb{P}\left\{T\left|\mathbb{D}_T\left(\gamma\right)\right| > \eta|\gamma - \gamma_0|T^{1-2\varphi} + m^2 \text{ for some } \gamma\right\} \\
&\leq \sum_{\ell=2}^{\infty}\mathbb{P}\left\{T\left|\mathbb{D}_T\left(\gamma\right)\right| > \eta(\ell - 1) + m^2 \text{ for some } \gamma \in A_{T\ell}\right\} \\
&\leq C'\sum_{\ell=2}^{\infty}\frac{\ell^2}{\left(\eta(\ell-1) + m^2\right)^4},
\end{aligned}$$

where the last equality is due to Lemma H.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 H.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

$$R_T^2 = \widetilde{n}^{2/3} \sup_{T^{-1+2\varphi} < |\gamma - \gamma_0| \leq C} \left[\left|\mathbb{D}_T\left(\gamma\right)\right| - \eta\left|\gamma - \gamma_0\right|_2^2\right],$$

where $\mathbb{D}_T\left(\gamma\right) = \frac{1}{\sqrt{N}T^{1-\varphi}}\sum_{t=1}^{T}\left(z_t\left(1_t\left(\gamma\right) - 1_t\left(\gamma_0\right)\right) - \mathbb{E}z_t\left(1_t\left(\gamma\right) - 1_t\left(\gamma_0\right)\right)\right)$. Then, for any $\varepsilon > 0$, we can find m such that

$$\begin{aligned}
\mathbb{P}\left\{R_T > m\right\} &= \mathbb{P}\left\{\widetilde{n}^{2/3}\left|\mathbb{D}_T\left(\gamma\right)\right| > \eta\widetilde{n}^{2/3}|\gamma - \gamma_0|^2 + m^2 \text{ for some } \gamma\right\} \\
&\leq \sum_{\ell=2}^{\infty}\mathbb{P}\left\{\widetilde{n}^{2/3}\left|\mathbb{D}_T\left(\gamma\right)\right| > \eta(\ell-1) + m^2 \text{ for some } \gamma \in A_{T\ell}\right\} \\
&\leq C'\sum_{\ell=2}^{\infty}\frac{\widetilde{n}^{2/3}}{\left(\eta(\ell-1) + m^2\right)^4}\frac{\ell}{\widetilde{n}^{2/3}} \leq \varepsilon
\end{aligned}$$

where the first and second inequalities follow from the union bound and Lemma H.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 H.4. First, we show the stochastic equicontinuity of $\mathbb{G}_n(f_n(z))$. For any positive ε and η, there exist $\delta > 0$ and N such that for all $n > N$,

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{|z-z'|_2<\delta}\left|\mathbb{G}_n(f_n(z)) - \mathbb{G}_n\left(f_n\left(z'\right)\right)\right|_2 > \eta\right\} \\
&\leq \mathbb{P}\Bigg\{\sup_{|z-z'|_2<\delta}\left|\mathbb{G}_n(f_n(z)) - \mathbb{G}_n\left(f_n\left(z'\right)\right)\right|_2 > \eta \text{ and } \left|f_n(z) - f_n\left(z'\right)\right|_2 \leq C\left|z-z'\right|_2 \\
&\qquad \text{and } \sup_z |f_n(z)|_2 \leq C\Bigg\} \\
&\quad + \mathbb{P}\left\{\left|f_n(z) - f_n\left(z'\right)\right|_2 > C\left|z-z'\right|_2\right\} + \mathbb{P}\left\{\sup_z |f_n(z)|_2 > C\right\} \\
&\leq \mathbb{P}\left\{\sup_{|x-x'|_2<\delta/C}\left|\mathbb{G}_n(x) - \mathbb{G}_n\left(x'\right)\right|_2 > \eta\right\} + \frac{\varepsilon}{2} \\
&\leq \varepsilon,
\end{aligned}$$

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

$$\mathbb{G}_n(f_n(z)) - \mathbb{G}_n(f(z)) \xrightarrow{p} 0$$

due to the stochastic equicontinuity of $\mathbb{G}_n$ as $f_n(z) \xrightarrow{p} f(z)$. Therefore, for any finite collection $(z_1, ..., z_p)'$, $(\mathbb{G}_n(f_n(z_1)), ..., \mathbb{G}_n(f_n(z_p)))' = (\mathbb{G}_n(f(z_1)), ..., \mathbb{G}_n(f(z_p)))' + o_P(1) \xrightarrow{d} (\mathbb{G}(f(z_1)), ..., \mathbb{G}(f(z_p)))'$ due to the weak convergence of $\mathbb{G}_n$. ∎